

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 36.5039 Seconds
(without alignments)
2747.774 Million cell updates/sec

Title: US-09-597-796C-4
Perfect score: 1802
Sequence: 1 MNSRRSRSLRWLSVLA.....QTKSGTGTGNTVLAEGPPA 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1802	100.0	355	2 AAW32367	AAW32367 Mycobacte
2	1802	100.0	355	2 AAW32435	AAW32435 Mycobacte
3	1802	100.0	355	2 AAW64307	AAW64307 Mycobacte
4	1802	100.0	355	2 AAW81670	AAW81670 M. tuberc
5	1802	100.0	355	2 AAY38972	AAY38972 M. tuberc
6	1802	100.0	355	2 AAY39109	AAY39109 M. tuberc
7	1802	100.0	355	4 AAU01890	AAU01890 M. tuberc
8	1802	100.0	355	5 AAE29701	AAE29701 Mycobacte
9	1802	100.0	355	5 AAE17565	AAE17565 Mycobacte
10	1792	99.4	355	2 AAY05000	AAY05000 Mycobacte
11	1792	99.4	355	4 AAO21137	AAO21137 Mycobacte
12	1792	99.4	355	4 AAG81110	AAG81110 Mycobacte
13	1792	99.4	379	2 AAY04830	AAY04830 Mycobacte
14	1634	90.7	330	5 AAE29702	AAE29702 Mycobacte
15	1634	90.7	330	5 AAE17566	AAE17566 Mycobacte
16	1631	90.5	330	5 AAE29703	AAE29703 Mycobacte
17	1631	90.5	330	5 AAE17567	AAE17567 Mycobacte
18	1631	90.5	330	7 ADA26372	ADA26372 Mycobacte
19	1631	90.5	723	7 ADA26354	ADA26354 Mycobacte
20	1631	90.5	1010	7 ADA26356	ADA26356 Mycobacte
21	1627	90.3	330	7 ADA26371	ADA26371 Mycobacte
22	999.5	55.5	1016	7 ADA26370	ADA26370 M. bovis
23	992	55.0	1154	7 ADA26368	ADA26368 Mycobacte
24	990.5	55.0	825	7 ADA26366	ADA26366 Mycobacte
25	989.5	54.9	813	7 ADA26367	ADA26367 Mycobacte

26	988	54.8	1022	7 ADA26369	ADA26369 Mycobacte
27	985	54.7	195	5 AAE29704	AAE29704 Mycobacte
28	985	54.7	195	5 AAE17568	AAE17568 Mycobacte
29	983.5	54.6	875	7 ADA26365	ADA26365 Mycobacte
30	983	54.6	596	2 AAY32070	AAY32070 Mycobacte
31	983	54.6	596	5 AAE29710	AAE29710 Mycobacte
32	983	54.6	596	5 AAE17574	AAE17574 Mycobacte
33	983	54.6	599	5 AAE17579	AAE17579 Antigenic
34	983	54.6	729	4 AAO22142	AAO22142 Ra12-H9-3
35	983	54.6	729	5 AAE29708	AAE29708 Mycobacte
36	983	54.6	729	5 AAE17572	AAE17572 Mycobacte
37	983	54.6	729	7 ADA26373	ADA26373 Mycobacte
38	983	54.6	930	5 AAE29731	AAE29731 Mycobacte
39	983	54.6	930	7 ADA26364	ADA26364 Mycobacte
40	980	54.4	729	5 AAE29709	AAE29709 Mycobacte
41	980	54.4	729	7 AAE17573	AAE17573 Mycobacte
42	980	54.4	729	7 ADA26374	ADA26374 Mycobacte
43	978	54.3	726	5 AAY74588	AAY74588 Antigenic
44	978	54.3	729	2 AAY32059	AAY32059 Mycobacte
45	676	37.5	132	2 AAW32354	AAW32354 Mycobacte

ALIGNMENTS

RESULT 1

AAW32367
ID AAW32367 standard; protein; 355 AA.

XX
AC AAW32367;
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TBra35.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-MAR-1996; 96US-00620280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
PA (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX Vedvick TH, Twardzik DR;
XX WPI; 1997-192904/17.
XX N-PSDB; AAT91414.
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
XX useful for diagnosis of M. tuberculosis infection.
XX
XX Example 3; Page 124-126; 190pp; English.

A new immunogenic polypeptide has been developed comprising an immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its variant differing only in conservative substitutions and/or modifications). The present sequence represents a M. tuberculosis antigen, TBra35. The immunogenic polypeptide can be used to diagnose M. tuberculosis infection by forming complexes with specific antibodies in the sample. Fragments of DNA encoding the immunogenic polypeptide can be used as diagnostic primers or probes and agents that bind to the antigen,

CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

XX Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSLAAVGLGLATAPAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
 DB 1 MNSRRSLRWSLLSLAAVGLGLATAPAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
 DB 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
 QY 121 YRTQDVAVLQLRGAGLPSAIGGVAVGEPVWAMGNSGGGQTPRAVPGRVVAGQTV 180
 DB 121 YRTQDVAVLQLRGAGLPSAIGGVAVGEPVWAMGNSGGGQTPRAVPGRVVAGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPDSGPPVNGLGQVVGNTAASDNFQLSQGGQGFA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPDSGPPVNGLGQVVGNTAASDNFQLSQGGQGFA 240
 QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVVGSAPAAASLGIST 300
 DB 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVVGSAPAAASLGIST 300
 QY 301 GGVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GGVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

AAW32435
 ID AAW32435 standard; protein; 355 AA.

XX AAW32435;

DT 08-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen TBra35.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

OS WO9709428-A2.

PN 13-MAR-1997.

XX 30-AUG-1996; 96WO-US014674.

PR 01-SEP-1995; 95US-00523436.

PR 22-SEP-1995; 95US-00533634.

PR 22-MAR-1996; 96US-00620874.

PR 05-JUN-1996; 96US-00659683.

PR 12-JUL-1996; 96US-00680574.

XX (CORI-) CORIXA CORP.

PA Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;

PI Vedvick TH, Twardzik DR;

XX WPI; 1997-192903/17.

DR N-PSDB; AAT91477.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.

PS

Example 3; Page 114-116; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TBra35. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)

XX Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSLAAVGLGLATAPAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
 DB 1 MNSRRSLRWSLLSLAAVGLGLATAPAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
 DB 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDVVG 120
 QY 121 YRTQDVAVLQLRGAGLPSAIGGVAVGEPVWAMGNSGGGQTPRAVPGRVVAGQTV 180
 DB 121 YRTQDVAVLQLRGAGLPSAIGGVAVGEPVWAMGNSGGGQTPRAVPGRVVAGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPDSGPPVNGLGQVVGNTAASDNFQLSQGGQGFA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPDSGPPVNGLGQVVGNTAASDNFQLSQGGQGFA 240
 QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVVGSAPAAASLGIST 300
 DB 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVVGSAPAAASLGIST 300
 QY 301 GGVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GGVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

AAW64307
 ID AAW64307 standard; protein; 355 AA.

XX AAW64307;

DT 17-OCT-2003 (revised)

DT 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen TBra35.

DE Tuberculosis; infection; diagnosis; antigen; TBra35.

XX Mycobacterium tuberculosis; strain H37Ra.

OS WO9816645-A2.

PN 23-APR-1998.

XX 07-OCT-1997; 97WO-US018214.

XX 11-OCT-1996; 96US-00729622.

PR 13-MAR-1997; 97US-00818111.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

PI Vedvick TS, Twardzik DR, Lodes MJ;

XX WPI; 1998-251292/22.

DR N-PSDB; AAV44355.
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
PT develop products for the detection of M. tuberculosis infection and
PT diagnosis of tuberculosis.
XX
XX
XX Example 3; Page 115-116; 250pp; English.
PS
XX This polypeptide comprises Mycobacterium tuberculosis soluble antigen
CC Tbra35. It is encoded by a DNA sequence (see AAV44355) isolated from a M.
CC tuberculosis strain H37Ra expression library with rabbit anti-sera raised
CC against M. tuberculosis supernatant. No significant homology was found
CC between Tbra35 and Genbank database sequences. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of a
CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic kits for
CC detecting M. tuberculosis infection in a patient using the above
CC polypeptides, antibodies or oligonucleotide probes and primers, for the
CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS
CC field)
XX
XX SQ Sequence 355 AA;
Query Match 100.0%; Score 1802; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNSRRSLRWLLSVLAAGLGLATAPACAPALSDQRFADFPALPLDPSAMVAQVA 60
Db 1 MNSRRSLRWLLSVLAAGLGLATAPACAPALSDQRFADFPALPLDPSAMVAQVA 60
Qy 61 PQVNTKLYNNAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDDVG 120
Db 61 PQVNTKLYNNAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDDVG 120
Qy 121 YDRTQDVAVLQIRGAGLPSAIIIGGVAVGEPVAMNGSGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLQIRGAGLPSAIIIGGVAVGEPVAMNGSGGGTTPRAVGRVVALGQTV 180
Qy 181 QASDLSLTGAETLNGLIQFDAAIQDSDGPGVNVGLGVGMNTAASDNFQSGGGGFA 240
Db 181 QASDLSLTGAETLNGLIQFDAAIQDSDGPGVNVGLGVGMNTAASDNFQSGGGGFA 240
Qy 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFLGLGVDDNNGARVORVVGSAAPASLGIST 300
Db 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFLGLGVDDNNGARVORVVGSAAPASLGIST 300
Qy 301 GDVITAVDGAPINSATAMADALNGHHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAPINSATAMADALNGHHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
RESULT 4
AAW81670
ID AAW81670 standard; protein; 355 AA.
XX
XX AAW81670;
XX
XX 27-JAN-1999 (first entry)
XX
XX M. tuberculosis immunogenic polypeptide Tbra35.
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9816646-A2.
XX
XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018293.
XX
XX 11-OCT-1996; 96US-00730510.
PR 13-MAR-1997; 97US-00818112.
XX
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedwick TS, Twardzik DR, Lodes MJ;
XX
XX WPI: 1998-261042/23.
DR N-PSDB; AAV64463.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
PT develop products for the detection of M. tuberculosis infection and for
PT diagnosis, treatment and prevention of tuberculosis.
XX
XX Example 3; Page 110-111; 230pp; English.
PS
XX This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
CC inducing protective immunity against tuberculosis (TB). This sequence can
CC be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis
XX
XX SQ Sequence 355 AA;
Query Match 100.0%; Score 1802; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNSRRSLRWLLSVLAAGLGLATAPACAPALSDQRFADFPALPLDPSAMVAQVA 60
Db 1 MNSRRSLRWLLSVLAAGLGLATAPACAPALSDQRFADFPALPLDPSAMVAQVA 60
Qy 61 PQVNTKLYNNAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDDVG 120
Db 61 PQVNTKLYNNAGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDDVG 120
Qy 121 YDRTQDVAVLQIRGAGLPSAIIIGGVAVGEPVAMNGSGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLQIRGAGLPSAIIIGGVAVGEPVAMNGSGGGTTPRAVGRVVALGQTV 180
Qy 181 QASDLSLTGAETLNGLIQFDAAIQDSDGPGVNVGLGVGMNTAASDNFQSGGGGFA 240
Db 181 QASDLSLTGAETLNGLIQFDAAIQDSDGPGVNVGLGVGMNTAASDNFQSGGGGFA 240
Qy 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFLGLGVDDNNGARVORVVGSAAPASLGIST 300
Db 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFLGLGVDDNNGARVORVVGSAAPASLGIST 300
Qy 301 GDVITAVDGAPINSATAMADALNGHHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAPINSATAMADALNGHHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
RESULT 5
AAV38972
ID AAV38972 standard; protein; 355 AA.
XX
XX AAV38972;
XX
XX 05-NOV-1999 (first entry)
XX
XX M. tuberculosis recombinant antigen protein Tbra35.
XX Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX
XX Mycobacterium tuberculosis.
XX
XX OS

```

PN WO9942118-A2.
XX
XX PD 26-AUG-1999.
XX
XX PF 17-FEB-1999; 99WO-US003265.
XX
XX PR 18-FEB-1998; 98US-00024753.
XX
XX PR 03-MAY-1998; 98US-00072596.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TS, Twardzik DR, Lodes MU, Hendrickson RC;
XX
XX DR WPI; 1999-527416/44.
XX DR N-PSDB; AA219053.
XX
XX PT New polypeptide comprising antigenic portions of M. tuberculosis.
XX
XX PS Example 3; Page 151-153; 323pp; English.
XX
XX CC This invention describes novel recombinant antigens and their encoding
XX CC nucleic acids derived from Mycobacterium tuberculosis. The novel
XX CC polypeptides are useful for detecting M. tuberculosis infection in a
XX CC biological sample by detecting antibodies which bind with the
XX CC polypeptides, and are useful as vaccines for immunizing against M.
XX CC tuberculosis infection. The new detection methods are needed as current
XX CC vaccination strategies do not provide 100% immunity
XX
XX SQ Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQA 60
DB 1 MNSRRSLRWLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQA 60

QY 61 PQVNTKLGYNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTGYVDVVG 120
DB 61 PQVNTKLGYNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTGYVDVVG 120

QY 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPRAVPGRVVVALGTV 180
DB 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPRAVPGRVVVALGTV 180

QY 181 QASDSLTAETLNGLIQFDDAIQPGDSGGPVVNGLGQVWGMNTAASDNFOLSGGGGFA 240
DB 181 QASDSLTAETLNGLIQFDDAIQPGDSGGPVVNGLGQVWGMNTAASDNFOLSGGGGFA 240

QY 241 IPIGQAMATAGIIRSGGSPVTHIGPTAFGLGVVDNNGARVQVVVGSAPAASLGIST 300
DB 241 IPIGQAMATAGIIRSGGSPVTHIGPTAFGLGVVDNNGARVQVVVGSAPAASLGIST 300

QY 301 GDVITAVDGAIPNSATAMADALNGHHHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAIPNSATAMADALNGHHHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 6
AA39109
ID AA39109 standard; protein; 355 AA.
XX
XX AC AA39109;
XX
XX DT 05-NOV-1999 (first entry)
XX
XX DE M. tuberculosis antigen TBa35 amino acid sequence.
XX
XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;
XX KW immune response; skin test.

```

```

XX Mycobacterium tuberculosis.
XX
XX OS WO9942076-A2.
XX
XX PN 26-AUG-1999.
XX
XX PD 17-FEB-1999; 99WO-US003268.
XX
XX PF 18-FEB-1998; 98US-00025197.
XX
XX PR 05-MAY-1998; 98US-00072967.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TS, Twardzik DR, Lodes MU, Hendrickson RC;
XX
XX DR WPI; 1999-527409/44.
XX DR N-PSDB; AA219265.
XX
XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin
XX PT tests and protective or therapeutic vaccines or compositions.
XX
XX PS Example 3; Page 108-109; 299pp; English.
XX
XX CC The present invention describes polypeptides comprising an immunogenic
XX CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
XX CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
XX CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
XX CC polypeptides fragments, can be used in pharmaceutical compositions or
XX CC vaccines to generate a protective or therapeutic immune response to M.
XX CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
XX CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
XX CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
XX CC to AA219460 and AA219083 to AA219225 are used in the exemplification of
XX CC the present invention
XX
XX SQ Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQA 60
DB 1 MNSRRSLRWLLSVLAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQA 60

QY 61 PQVNTKLGYNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTGYVDVVG 120
DB 61 PQVNTKLGYNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTGYVDVVG 120

QY 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPRAVPGRVVVALGTV 180
DB 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPRAVPGRVVVALGTV 180

QY 181 QASDSLTAETLNGLIQFDDAIQPGDSGGPVVNGLGQVWGMNTAASDNFOLSGGGGFA 240
DB 181 QASDSLTAETLNGLIQFDDAIQPGDSGGPVVNGLGQVWGMNTAASDNFOLSGGGGFA 240

QY 241 IPIGQAMATAGIIRSGGSPVTHIGPTAFGLGVVDNNGARVQVVVGSAPAASLGIST 300
DB 241 IPIGQAMATAGIIRSGGSPVTHIGPTAFGLGVVDNNGARVQVVVGSAPAASLGIST 300

QY 301 GDVITAVDGAIPNSATAMADALNGHHHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAIPNSATAMADALNGHHHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 7
AAU01890
ID AAU01890 standard; protein; 355 AA.
XX
XX AC AAU01890;

```

XX 29-AUG-2001 (first entry)
 XX M. tuberculosis antigen Tba35 (Mtb32A).
 DE Tba35; Mtb32A; antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease.
 XX Mycobacterium tuberculosis.
 OS WO200124820-A1.
 XX 12-APR-2001.
 XX 10-OCT-2000; 2000WO-US028095.
 XX 07-OCT-1999; 99US-0156338P.
 PR 07-OCT-1999; 99US-0156425P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
 DR WPI; 2001-290576/30.
 XX N-PSDB; AAS03781.
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens.
 XX
 XX Disclosure; Page 153-154; 168pp; English.
 XX The sequence represents Mycobacterium tuberculosis Tba35 (also known as
 CC Mtb32A), an M. tuberculosis antigen. Compositions comprising at least 2
 CC heterologous antigens, as a fusion protein, and vectors expressing the
 CC fusion proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS
 XX
 XX Sequence 355 AA;
 XX
 XX Query Match 100.0%; Score 1802; DB 4; Length 355;
 XX Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 XX Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSNRRRLRWLLSVLAAVGLGLATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNRRRLRWLLSVLAAVGLGLATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNVNITKLGYNNAVAGGTGIVIDPVGVLNNHVIAGATDINAFSGSGQTYGVDDVVG 120
 DB 61 PQVNVNITKLGYNNAVAGGTGIVIDPVGVLNNHVIAGATDINAFSGSGQTYGVDDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGSGGPPVNGLGQVVMGNTAASDNFQLSQGGGGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGSGGPPVNGLGQVVMGNTAASDNFQLSQGGGGA 240
 QY 241 IPIGQMAIAGQIRSGGSPVTHIGPTAFLGLGVDDNNGGARVQRVVGSAASLGIST 300
 DB 241 IPIGQMAIAGQIRSGGSPVTHIGPTAFLGLGVDDNNGGARVQRVVGSAASLGIST 300
 QY 301 GDVITAVDGAIPNSATAMADALNGHPDVISVNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAIPNSATAMADALNGHPDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 8

AAE29701
 ID AAE29701 standard; protein; 355 AA.
 XX
 AC AAE29701;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium tuberculosis Mtb32A antigenic protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; Mtb32A antigen;
 KW Ra35FL.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200272792-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 13-MAR-2002; 2002WO-US008223.
 XX
 PR 13-MAR-2001; 2001US-0275837P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Brannon M, Guderian J;
 XX
 DR WPI; 2002-759844/82.
 XX
 DR N-PSDB; AAD47076.
 XX
 PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 XX tuberculosis.
 XX
 PS Disclosure; Page 78-79; 155pp; English.
 XX
 CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis Mtb32A antigenic protein. Mtb32A is also referred to as
 CC Ra35FL
 XX
 XX Sequence 355 AA;
 XX
 XX Query Match 100.0%; Score 1802; DB 5; Length 355;
 XX Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 XX Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSNRRRLRWLLSVLAAVGLGLATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNRRRLRWLLSVLAAVGLGLATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNVNITKLGYNNAVAGGTGIVIDPVGVLNNHVIAGATDINAFSGSGQTYGVDDVVG 120
 DB 61 PQVNVNITKLGYNNAVAGGTGIVIDPVGVLNNHVIAGATDINAFSGSGQTYGVDDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGSGGPPVNGLGQVVMGNTAASDNFQLSQGGGGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGSGGPPVNGLGQVVMGNTAASDNFQLSQGGGGA 240

QY 241 IPIQWATAGAIRSGGGSPVTHIGPTAFGLGVVDNNGARVQVVGSAAPASLGIST 300
 DB 241 IPIQWATAGAIRSGGGSPVTHIGPTAFGLGVVDNNGARVQVVGSAAPASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 9

AAE17565
 ID AAE17565 standard; protein; 355 AA.

XX
 AC AAE17565;

DT 22-APR-2002 (first entry)

XX Mycobacterium species MTB32A (Ra35FL) protein.

DE Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.
 XX
 OS Mycobacterium sp.
 XX
 FN WO200198460-A2.
 XX
 PD 27-DEC-2001.
 XX
 XX 20-JUN-2001; 2001WO-US019959.
 PF
 XX 20-JUN-2000; 2000US-00597796.
 PR
 PR 01-FEB-2001; 2001US-0265737P.
 XX
 XX (CORI-) CORIXA CORP.
 PA

PI Skeiky Y, Reed S, Alderson M;

XX WPI; 2002-147798/19.
 DR
 DR N-PSDB; AAD28335.
 XX

PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.

PS Claim 70; Page 94-95; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) protein
 XX
 SQ Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 5; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNRRRLSWLLSVLAAYVGLGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNRRRLSWLLSVLAAYVGLGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNVINTKLGYNNAVAGAGTIVIDPNGVLTNNHVIAGATDINAFSVSGQTYGVDDVVG 120
 DB 61 PQVNVINTKLGYNNAVAGAGTIVIDPNGVLTNNHVIAGATDINAFSVSGQTYGVDDVVG 120
 QY 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGCEPVVAMNGSGGGTTPRAYPGRVVALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGCEPVVAMNGSGGGTTPRAYPGRVVALGQTV 180
 QY 181 QASDSLGTGABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMTAASDNFQLSQGGQGFA 240
 DB 181 QASDSLGTGABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMTAASDNFQLSQGGQGFA 240
 QY 241 IPIQWATAGAIRSGGGSPVTHIGPTAFGLGVVDNNGARVQVVGSAAPASLGIST 300
 DB 241 IPIQWATAGAIRSGGGSPVTHIGPTAFGLGVVDNNGARVQVVGSAAPASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 10

AAV05000

ID AAV05000 standard; protein; 355 AA.

XX
 AC AAV05000;

DT 06-JUL-1999 (first entry)

XX Mycobacterium species protein sequence 50D.

DE Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 OS Mycobacterium sp.
 XX
 DN WO9909186-A2.
 PD
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR001813.
 XX
 PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX
 PA (INSP) INST PASTEUR.
 XX

PI Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;

PI Coguet De La Salmoniere Y;

XX WPI; 1999-181045/15.
 DR
 DR N-PSDB; AAX34251.
 XX

PT Mycobacterial DNA vectors containing reporter constructs - for

PT identifying coding or promoter sequences involved in infection-associated

PT protein expression.

XX
 PS Claim 32; Fig 50D; 309pp; French.
 XX

CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection

SQ Sequence 355 AA;

Query Match 99.4%; Score 1792; DB 2; Length 355;

Best Local Similarity 99.4%; Pred. No. 4e-125;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWLLSVLAAGVGLATAPAAQAPALSDQRFADPPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWLLSVLAAGVGLATAPAAQAPALSDQRFADPPALPLDPSAMVAQV 60
QY 61 PQVNTKLGYNNAVGAGTGVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
DB 61 PQVNTKLGYNNAVGAGTGVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
QY 121 YRTOQDAVQLRGAGGLPSAAIGGVAVGEPVAVMGNSGGGTPRAVGRVVALGQTV 180
DB 121 YRTOQDAVQLRGAGGLPSAAIGGVAVGEPVAVMGNSGGGTPRAVGRVVALGQTV 180
QY 181 QASDSLTAETLNGLIQFDAAIQGDSGGFPVNGLGQVGMNTAASDNFOLSGGGGFA 240
DB 181 QASDSLTAETLNGLIQFDAAIQGDSGGFPVNGLGQVGMNTAASDNFOLSGGGGFA 240
QY 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGARVQVVGSAPAASLGIST 300
DB 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGARVQVVGSAPAASLGIST 300
QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 11
AAO22137
ID AAO22137 standard; protein; 355 AA.
XX AAO22137;
XX AC
XX AC
DT 03-OCT-2002 (first entry)
XX Mycobacterium tuberculosis MTB32A protein.
XX DE
XX Ral2; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
XX KW vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
XX KW immunogen; cytokine.
XX OS
XX Mycobacterium tuberculosis.
XX PN
XX WO200125401-A2.
XX PD
XX 12-APR-2001.
XX PF
XX 06-OCT-2000; 2000WO-US027652.
XX PR
XX 07-OCT-1999; 99US-0158585P.
XX PA (CORI-) CORIXA CORP.
XX PI
XX Skeiky Y, Guderian J;
XX PR
XX WPI; 2001-266299/27.
XX DR N-PSDB; AAL40768.
XX XX
XX PT Recombinant nucleic acid molecule for producing high yield expression of
XX PT desired fusion polypeptides, encodes fusion polypeptide comprising
XX PT Mycobacterium tuberculosis coding sequence and heterologous polypeptide.
XX PS
XX Disclosure; Fig 1; 39pp; English.
XX XX
XX The invention relates to a recombinant nucleic acid molecule encoding a
XX CC fusion polypeptide, comprising a polynucleotide sequence of Ral2, a 14
XX CC kDa C-terminal fragment of serine protease antigen MTB32A of
XX CC Mycobacterium tuberculosis, and a heterologous polynucleotide sequence.
XX CC The recombinant fusion nucleic acids and polypeptides are useful for
XX CC providing stable and high yield expression of fusion polypeptides of both
XX CC eukaryotic and prokaryotic origin and to encode a protein product for use
XX CC as an antigen for detecting serum antibodies. The presence of serum

CC antibodies to M. tuberculosis antigens in an individual indicates that
CC the individual is infected with it. The fusion polypeptides are useful as
CC sources of proteins for monitoring binding of serum antibodies to fusion
CC proteins and as an immunogen to induce and/or enhance immune responses.
CC The coding sequences can be ligated with a coding sequence of another
CC molecule such as M. tuberculosis antigens, a cytokine or an adjuvant and
CC can be used in vivo as a DNA vaccine. This sequence represents the
CC Mycobacterium tuberculosis MTB32A protein
XX
SQ Sequence 355 AA;
Query Match 99.4%; Score 1792; DB 4; Length 355;
Best Local Similarity 99.4%; Pred. No. 4e-125;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWLLSVLAAGVGLATAPAAQAPALSDQRFADPPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWLLSVLAAGVGLATAPAAQAPALSDQRFADPPALPLDPSAMVAQV 60
QY 61 PQVNTKLGYNNAVGAGTGVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
DB 61 PQVNTKLGYNNAVGAGTGVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
QY 121 YRTOQDAVQLRGAGGLPSAAIGGVAVGEPVAVMGNSGGGTPRAVGRVVALGQTV 180
DB 121 YRTOQDAVQLRGAGGLPSAAIGGVAVGEPVAVMGNSGGGTPRAVGRVVALGQTV 180
QY 181 QASDSLTAETLNGLIQFDAAIQGDSGGFPVNGLGQVGMNTAASDNFOLSGGGGFA 240
DB 181 QASDSLTAETLNGLIQFDAAIQGDSGGFPVNGLGQVGMNTAASDNFOLSGGGGFA 240
QY 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGARVQVVGSAPAASLGIST 300
DB 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGARVQVVGSAPAASLGIST 300
QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 12
AAG81110
ID AAG81110 standard; protein; 355 AA.
XX AC
XX AC
XX AAG81110;
DT 04-SEP-2001 (first entry)
XX Mycobacterium tuberculosis potential drug target protein SEQ ID 161.
XX DE
XX Drug target; growth; organism viability; characterisation.
XX KW
XX Mycobacterium tuberculosis.
XX OS
XX WO200135317-A1.
XX PN
XX 17-MAY-2001.
XX PF
XX 13-NOV-2000; 2000WO-US031152.
XX XX
XX 12-NOV-1999; 99US-0165086P.
XX PR 12-NOV-1999; 99US-0165124P.
XX PR 01-FEB-2000; 2000US-0179531P.
XX XX
XX (REGC) UNIV CALIFORNIA.
XX PA
XX Eisenberg D, Rotstein SH, Marcotte EM;
XX WPI; 2001-329193/34.
XX DR N-PSDB; AAH51961.
XX XX
XX Identifying nucleotide or polypeptide sequence for use as drug target,
XX PT involves providing algorithm that analyzes a functional relationship

PT between nucleotide or polypeptide sequences, and comparing the sequences.
 XX Disclosure; Page 157; 207pp; English.
 PS This invention relates to a method for identifying a nucleotide or
 XX polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
 CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism
 XX Sequence 355 AA;

Query Match 99.4%; Score 1792; DB 4; Length 355;
 Best Local Similarity 99.4%; Pred. No. 4e-125;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 Db 1 MNSRRSLRWWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVG 60
 QY 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 Db 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 QY 121 YDRTQDVAVQLRGAGGLPSAAIGGVAVGEPVWAMNSGGGGTTPRAVPGRVVALGQTV 180
 Db 121 YDRTQDVAVQLRGAGGLPSAAIGGVAVGEPVWAMNSGGGGTTPRAVPGRVVALGQTV 180
 QY 181 QASDSLTCABETLNGLIQFDAAIQDSDGGPVVNGLGQVWGMNTAASDNFOLSQGGQGFA 240
 Db 181 QASDSLTCABETLNGLIQFDAAIQDSDGGPVVNGLGQVWGMNTAASDNFOLSQGGQGFA 240
 QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVORVVGSAAPASLGIST 300
 Db 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVORVVGSAAPASLGIST 300
 QY 301 GDVITAVDGPINSATAMADALNGHPGDVIVSVWQTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGPINSATAMADALNGHPGDVIVSVWQTKSGGTRTGNVTLAEGPPA 355

RESULT 13
 AA04830 standard; protein; 379 AA.
 XX AA04830;
 AC
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species protein sequence 50F.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 OS Mycobacterium sp.
 XX
 PN WO990186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR001813.
 XX
 PR 14-AUG-1997; 97FR-00010404.
 PR
 XX 11-SEP-1997; 97FR-00011325.
 XX

PA (INSP) INST PASTEUR.
 XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 XX WPI; 1999-181045/15.
 DR N-PSDB; AAX34252.
 XX
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX
 PS Claim 32; Fig 50F; 309pp; French.
 XX
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 SQ Sequence 379 AA;

Query Match 99.4%; Score 1792; DB 2; Length 379;
 Best Local Similarity 99.4%; Pred. No. 4.3e-125;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 Db 25 MNSRRSLRWWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVG 84
 QY 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 Db 85 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 144
 QY 121 YDRTQDVAVQLRGAGGLPSAAIGGVAVGEPVWAMNSGGGGTTPRAVPGRVVALGQTV 180
 Db 145 YDRTQDVAVQLRGAGGLPSAAIGGVAVGEPVWAMNSGGGGTTPRAVPGRVVALGQTV 204
 QY 181 QASDSLTCABETLNGLIQFDAAIQDSDGGPVVNGLGQVWGMNTAASDNFOLSQGGQGFA 240
 Db 205 QASDSLTCABETLNGLIQFDAAIQDSDGGPVVNGLGQVWGMNTAASDNFOLSQGGQGFA 264
 QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVORVVGSAAPASLGIST 300
 Db 265 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVORVVGSAAPASLGIST 324
 QY 301 GDVITAVDGPINSATAMADALNGHPGDVIVSVWQTKSGGTRTGNVTLAEGPPA 355
 Db 325 GDVITAVDGPINSATAMADALNGHPGDVIVSVWQTKSGGTRTGNVTLAEGPPA 379

RESULT 14
 AA29702 standard; protein; 330 AA.
 ID AA29702 standard; protein; 330 AA.
 XX
 AC AA29702;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium tuberculosis mature Ra35 antigenic protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; Ra35 antigen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 182 /note= "Encoded by GAG"
 FT Misc-difference 183 /note= "Encoded by GCG"
 FT
 XX WO200272792-A2.
 PN

XX 19-SEP-2002.
 XX 13-MAR-2002; 2002WO-US008223.
 XX 13-MAR-2001; 2001US-0275837P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Brannon M, Guderian J;
 XX WPI; 2002-759844/82.
 XX N-PSDB; AAD47077.
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX Disclosure; Page 79-80; 155pp; English.
 XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis mature Ra35 (N-terminus of MTB32A; Ra35FL) antigenic protein
 XX Sequence 330 AA;
 SQ
 Query Match 90.7%; Score 1634; DB 5; Length 330;
 Best Local Similarity 99.4%; Pred. No. 2.1e-113;
 Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNNITKLGYNNAVAGTGIVIDPVGVL 92
 Db 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNNITKLGYNNAVAGTGIVIDPVGVL 67
 QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 152
 Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
 QY 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQDGSQGPV 212
 Db 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQDGSQGPV 187
 QY 213 VNLGQVVGNTAASDNFQSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGL 272
 Db 188 VNLGQVVGNTAASDNFQSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGL 247
 QY 273 GVVDNNGGARVQVRVVGSPAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 332
 Db 248 GVVDNNGGARVQVRVVGSPAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 307
 QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
 Db 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 RESULT 15
 AAEL17566
 ID AAEL17566 standard; protein; 330 AA.
 XX
 AC AAEL17566;
 XX
 DT 22-APR-2002 (first entry)

XX Mycobacterium sp. MTB32A (Ra35FL) mature protein.
 DE Fusion protein; antigen; serological sensitivity; immune response;
 XX tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.
 KW
 XX Mycobacterium sp.
 OS
 XX Key Location/Qualifiers
 FH Region 8..202
 FT /note= "Ra35 N-terminal peptide"
 FT Misc-difference 182 /note= "Encoded by GAG"
 FT /note= "Encoded by GCG"
 FT Misc-difference 183 /note= "Encoded by GCG"
 FT Region 199..330
 FT /note= "Ra35 C-terminal peptide, Ra12"
 XX
 PN WO200198460-A2.
 XX
 XX 27-DEC-2001.
 PD
 XX 20-JUN-2001; 2001WO-US019959.
 XX
 XX 20-JUN-2001; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0265737P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky Y, Reed S, Alderson M;
 FI WPI; 2002-147798/19.
 XX N-PSDB; AAD28336.
 DR
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 PT
 XX Claim 70; Fig 6; 136pp; English.
 PS
 XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) mature protein
 CC
 XX Sequence 330 AA;
 SQ
 Query Match 90.7%; Score 1634; DB 5; Length 330;
 Best Local Similarity 99.4%; Pred. No. 2.1e-113;
 Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNNITKLGYNNAVAGTGIVIDPVGVL 92
 Db 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNNITKLGYNNAVAGTGIVIDPVGVL 67
 QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 152
 Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
 QY 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQDGSQGPV 212
 Db 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQDGSQGPV 187
 QY 213 VNLGQVVGNTAASDNFQSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGL 272
 Db 188 VNLGQVVGNTAASDNFQSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGL 247
 QY 273 GVVDNNGGARVQVRVVGSPAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 332
 Db 248 GVVDNNGGARVQVRVVGSPAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 307
 QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
 Db 308 VTWQTKSGGTRTGNVTLAEGPPA 330

Db 68 NNHVIAGATDINAFSGGQTYGVYVGYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEP 127
Qy 153 VVAMNGSGGGCTPRAYPGRWVVALGQTVQASDSLTTGAETLNGLIQFDAAIQPGDSGGPV 212
Db 128 VVAMNGSGGGCTPRAYPGRWVVALGQTVQASDSLTTGAETLNGLIQFDAAIQPGDSGGPV 187
Qy 213 VVGLGQVVGMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGGSPVHIGPTAFLGL 272
Db 188 VVGLGQVVGMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGGSPVHIGPTAFLGL 247
Qy 273 GVVDNNGNGARVQVRVVGSAFASLIGISTGVDVITAVDGAIPINSATAMADALNGHHPGDVIS 332
Db 248 GVVDNNGNGARVQVRVVGSAFASLIGISTGVDVITAVDGAIPINSATAMADALNGHHPGDVIS 307
Qy 333 VVWQTKSGGTRTGNVTILAEGPPA 355
Db 308 VVWQTKSGGTRTGNVTILAEGPPA 330

Search completed: June 30, 2004, 16:48:46
Job time : 38.5039 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 10.7991 Seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796C-4
Perfect score: 1802
Sequence: 1 MNSRRSLRWSLWSVLA.....QTKSGTRGTGNTLAEGPPA 355
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgm2_6/ptodata/2/iaa/5A COMB.pap.*
2: /cgm2_6/ptodata/2/iaa/5B COMB.pap.*
3: /cgm2_6/ptodata/2/iaa/6A COMB.pap.*
4: /cgm2_6/ptodata/2/iaa/6B COMB.pap.*
5: /cgm2_6/ptodata/2/iaa/PTCTUS COMB.pap.*
6: /cgm2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1802	100.0	355	3	US-08-818-112-79
2	1802	100.0	355	4	US-08-818-111-80
3	1802	100.0	355	4	US-09-056-556-79
4	1802	100.0	355	4	US-09-072-596-80
5	1802	100.0	355	4	US-09-072-967-79
6	983	54.6	596	4	US-09-287-849-26
7	983	54.6	729	4	US-09-223-040-2
8	983	54.6	729	4	US-09-287-849-2
9	676	37.5	132	3	US-08-818-112-66
10	676	37.5	132	4	US-08-818-111-67
11	676	37.5	132	4	US-09-056-556-66
12	676	37.5	132	4	US-09-072-596-67
13	676	37.5	132	4	US-09-072-967-66
14	676	37.5	132	4	US-09-636-215-819
15	676	37.5	132	4	US-09-636-215-848
16	676	37.5	132	4	US-09-685-166A-819
17	676	37.5	132	4	US-09-685-166A-848
18	671	37.2	231	4	US-09-287-849-28
19	628	34.9	224	4	US-09-636-215-825
20	628	34.9	224	4	US-09-685-166A-825
21	628	34.9	267	4	US-09-643-597-352
22	628	34.9	267	4	US-09-606-421B-352
23	628	34.9	273	4	US-09-736-457-1864
24	628	34.9	299	4	US-09-643-597-354
25	628	34.9	299	4	US-09-606-421B-354
26	628	34.9	304	4	US-09-636-215-835
27	628	34.9	304	4	US-09-685-166A-835

Sequence 1863, App
Sequence 852, App
Sequence 852, App
Sequence 349, App
Sequence 349, App
Sequence 333, App
Sequence 333, App
Sequence 196, App
Sequence 196, App
Sequence 196, App
Sequence 353, App
Sequence 353, App
Sequence 337, App
Sequence 337, App
Sequence 309, App
Sequence 309, App
Sequence 325, App
Sequence 325, App

ALIGNMENTS

RESULT 1
US-08-818-112-79
; Sequence 79, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-79

Query Match 100.0%; Score 1802; DB 3; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNTKLGYNNAVAGAGTGIVIDPNGVVLNNHVIAGATDINAFSVSGGOTYGVDDVVG 120
 DB 61 PQVNTKLGYNNAVAGAGTGIVIDPNGVVLNNHVIAGATDINAFSVSGGOTYGVDDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPRVWALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPRVWALGQTV 180
 QY 181 QASDSLGTGAETLNGLIQFDDAIOQDSDGPPVNGLVGVVGMNTAASDNFQLSQGGGFA 240
 DB 181 QASDSLGTGAETLNGLIQFDDAIOQDSDGPPVNGLVGVVGMNTAASDNFQLSQGGGFA 240
 QY 241 IPTGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVORVWGSAPAAASLGIST 300
 DB 241 IPTGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVORVWGSAPAAASLGIST 300
 QY 301 GDVITAVDGPAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGPAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

US-08-818-111-80
 ; Sequence 80, Application US/08818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/08/818,111
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 80:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 355 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-818-111-80

Query Match 100.0%; Score 1802; DB 4; Length 355;
 Best Local Similarity 100.0%; Pred. No. 2.1e-151;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNTKLGYNNAVAGAGTGIVIDPNGVVLNNHVIAGATDINAFSVSGGOTYGVDDVVG 120
 DB 61 PQVNTKLGYNNAVAGAGTGIVIDPNGVVLNNHVIAGATDINAFSVSGGOTYGVDDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPRVWALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPRVWALGQTV 180
 QY 181 QASDSLGTGAETLNGLIQFDDAIOQDSDGPPVNGLVGVVGMNTAASDNFQLSQGGGFA 240
 DB 181 QASDSLGTGAETLNGLIQFDDAIOQDSDGPPVNGLVGVVGMNTAASDNFQLSQGGGFA 240
 QY 241 IPTGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVORVWGSAPAAASLGIST 300
 DB 241 IPTGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVORVWGSAPAAASLGIST 300
 QY 301 GDVITAVDGPAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGPAPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

US-09-056-556-79
 ; Sequence 79, Application US/09056556
 ; Patent No. 6350456
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 ; NUMBER OF SEQUENCES: 241
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/09/056,556
 ; FILING DATE: 07-APR-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.457
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 79:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 355 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-056-556-79

Query Match 100.0%; Score 1802; DB 4; Length 355;
 Best Local Similarity 100.0%; Pred. No. 2.1e-151;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60

Db 1 MNSRRSLRWSLLSVLAAYGLGATAPAAQAPALSDRFPADFPALPLDPSAMVAQVA 60
QY 61 PQVWINTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVG 120
Db 61 PQVWINTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVG 120
QY 121 YDRTQDVAVLQRGAGGLPSAAIGGVAVGPPVAMGNSGGGQTPRAVPGRVVALGQTV 180
Db 121 YDRTQDVAVLQRGAGGLPSAAIGGVAVGPPVAMGNSGGGQTPRAVPGRVVALGQTV 180
QY 181 QASDSLTAETLNGLIQFDDAIQPDGSGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
Db 181 QASDSLTAETLNGLIQFDDAIQPDGSGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQVVGSAAPAASLGIST 300
Db 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQVVGSAAPAASLGIST 300
QY 301 GDVITAVDGAIPNSATAMADALNGHHPGDVSVNMQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAIPNSATAMADALNGHHPGDVSVNMQTKSGGTRTGNVTLAEGPPA 355

RESULT 4

US-09-072-596-80
; Sequence 80, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-80

Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAYGLGATAPAAQAPALSDRFPADFPALPLDPSAMVAQVA 60
Db 1 MNSRRSLRWSLLSVLAAYGLGATAPAAQAPALSDRFPADFPALPLDPSAMVAQVA 60
QY 61 PQVWINTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVG 120
Db 61 PQVWINTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVG 120
QY 121 YDRTQDVAVLQRGAGGLPSAAIGGVAVGPPVAMGNSGGGQTPRAVPGRVVALGQTV 180
Db 121 YDRTQDVAVLQRGAGGLPSAAIGGVAVGPPVAMGNSGGGQTPRAVPGRVVALGQTV 180
QY 181 QASDSLTAETLNGLIQFDDAIQPDGSGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
Db 181 QASDSLTAETLNGLIQFDDAIQPDGSGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQVVGSAAPAASLGIST 300
Db 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQVVGSAAPAASLGIST 300
QY 301 GDVITAVDGAIPNSATAMADALNGHHPGDVSVNMQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAIPNSATAMADALNGHHPGDVSVNMQTKSGGTRTGNVTLAEGPPA 355

RESULT 5

US-09-072-967-79
; Sequence 79, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-79

```
Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151; Indels 0; Gaps 0;
Matches 355; Conservative 0; Mismatches 0;

QY 1 MNSRRSLRWSLLSLVAAGLGLATAPAAQAPALPSQDRFADFPALPLDPSAMVAQVA 60
Db 1 MNSRRSLRWSLLSLVAAGLGLATAPAAQAPALPSQDRFADFPALPLDPSAMVAQVA 60
QY 61 PQVNTNTKLYNNVAGCTGIVIDPQVLTNNHVIAGATDINAFSVGSGQTYGVVVG 120
Db 61 PQVNTNTKLYNNVAGCTGIVIDPQVLTNNHVIAGATDINAFSVGSGQTYGVVVG 120
QY 121 YDRTPQVAVLQRLGAGGLPSAALGGVAVGEPVPMVAMNSGGQGTTPRAVPGRWALGQTV 180
Db 121 YDRTPQVAVLQRLGAGGLPSAALGGVAVGEPVPMVAMNSGGQGTTPRAVPGRWALGQTV 180
QY 181 QASDSLTGAETLNGLIQFADAIQPDSCGPPVNGLGQVVGWNTAASDNFQLSGGQGFA 240
Db 181 QASDSLTGAETLNGLIQFADAIQPDSCGPPVNGLGQVVGWNTAASDNFQLSGGQGFA 240
QY 241 IPIQWALAGQIRSGGSGPTVHIGTAFGLGVVDNNGGARVORVVGSAAPASLGIST 300
Db 241 IPIQWALAGQIRSGGSGPTVHIGTAFGLGVVDNNGGARVORVVGSAAPASLGIST 300
QY 301 GDVITAVDGAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLASGPPA 355
Db 301 GDVITAVDGAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLASGPPA 355

RESULT 6
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 54.6%; Score 983; DB 4; Length 596;
Best Local Similarity 97.5%; Pred. No. 1.1e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSQDRFADFPALPLDPSAMVAQVAPQVWNTKLYNNVAGCTGIVIDPN 87
Db 397 AAGDIAPALSQDRFADFPALPLDPSAMVAQVAPQVWNTKLYNNVAGCTGIVIDPN 456

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2

Query Match 54.6%; Score 983; DB 4; Length 729;
Best Local Similarity 97.5%; Pred. No. 1.5e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSQDRFADFPALPLDPSAMVAQVAPQVWNTKLYNNVAGCTGIVIDPN 87
Db 530 AAGDIAPALSQDRFADFPALPLDPSAMVAQVAPQVWNTKLYNNVAGCTGIVIDPN 589
QY 88 GVLTTNNHVIAGATDINAFSVGSGQTYGVVVGVDRTQDVAVLQRLGAGGLPSAAIGGV 147
Db 590 GVLTTNNHVIAGATDINAFSVGSGQTYGVVVGVDRTQDVAVLQRLGAGGLPSAAIGGV 649
QY 148 AVGEPVPMVAMNSGGQGTTPRAVPGRWALGQTVQASDSLTGAETLNGLIQFADAIQPD 207
Db 650 AVGEPVPMVAMNSGGQGTTPRAVPGRWALGQTVQASDSLTGAETLNGLIQFADAIQPD 709
QY 208 SGGPVVNGLGQVVGWNTAAS 227
Db 710 SGGPVVNGLGQVVGWNTAAS 729

RESULT 8
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-223-040-2

Query Match 54.6%; Score 983; DB 4; Length 729;
Best Local Similarity 97.5%; Pred. No. 1.5e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSQDRFADFPALPLDPSAMVAQVAPQVWNTKLYNNVAGCTGIVIDPN 87
Db 530 AAGDIAPALSQDRFADFPALPLDPSAMVAQVAPQVWNTKLYNNVAGCTGIVIDPN 589
QY 88 GVLTTNNHVIAGATDINAFSVGSGQTYGVVVGVDRTQDVAVLQRLGAGGLPSAAIGGV 147
Db 590 GVLTTNNHVIAGATDINAFSVGSGQTYGVVVGVDRTQDVAVLQRLGAGGLPSAAIGGV 649
QY 148 AVGEPVPMVAMNSGGQGTTPRAVPGRWALGQTVQASDSLTGAETLNGLIQFADAIQPD 207
Db 650 AVGEPVPMVAMNSGGQGTTPRAVPGRWALGQTVQASDSLTGAETLNGLIQFADAIQPD 709
QY 208 SGGPVVNGLGQVVGWNTAAS 227
Db 710 SGGPVVNGLGQVVGWNTAAS 729

RESULT 8
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2
```

```

; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRN
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tr1-fusion
US-09-287-849-2

Query Match          54.6%; Score 983; DB 4; Length 729;
Best Local Similarity 97.5%; Pred. No. 1.5e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPALSDRFADFPALPLDPSAMVAQVQVNVNINIKLYNNVAGAGTGIVIDPN 87
DB 530 AAGDIAPALSDRFADFPALPLDPSAMVAQVQVNVNINIKLYNNVAGAGTGIVIDPN 589
QY 88 GVLITNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVIQLRGAGGLPSAAIGGV 147
DB 590 GVLITNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVIQLRGAGGLPSAAIGGV 649
QY 148 AVGEFVVMGNSGCGGTTPRAVGRVVALGQTVQASDLSLTCAEETLNGLIQFRAAIQPGD 207
DB 650 AVGEFVVMGNSGCGGTTPRAVGRVVALGQTVQASDLSLTCAEETLNGLIQFRAAIQPGD 709
QY 208 SGGFVWNGLGQVVGNTAAS 227
DB 710 SGGFVWNGLGQVVGNTAAS 729

RESULT 9
US-08-818-112-66
; Sequence 66, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids

Query Match          37.5%; Score 676; DB 3; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASNFOLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFLGLGVVDNNGNGAR 283
DB 1 TAASNFOLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFLGLGVVDNNGNGAR 60
QY 284 VORVVGSPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIVSNWQTSGGTR 343
DB 61 VORVVGSPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIVSNWQTSGGTR 120
QY 344 TGNVTLAEGPPA 355
DB 121 TGNVTLAEGPPA 132

RESULT 10
US-08-818-111-67
; Sequence 67, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
```



```

; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 819
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-636-215-819

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFOLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFLGLGVVDNNGNGAR 283
Ddb 1 TAASDNFOLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFLGLGVVDNNGNGAR 60

QY 284 VQRVGSAPAASLGISTGTVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
Ddb 61 VQRVGSAPAASLGISTGTVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Ddb 121 TGNVTLAEGPPA 132

RESULT 15
US-09-636-215-848
; Sequence 848, Application US/09636215
; Patent No. 8620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 848
; LENGTH: 132

```

TYPE: PRT
; ORGANISM: Homo sapiens
US-09-636-215-848

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	224	TAASDNFQLSQGGGPAIPIGQAMAIAGQIRSGGGSPTVHIGTAPLGLGVVDNNGGAR	283
Db	1	TAASDNFQLSQGGGPAIPIGQAMAIAGQIRSGGGSPTVHIGTAPLGLGVVDNNGGAR	60
QY	284	VQRVVGSAAPASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVISVNMQTKSGGTR	343
Db	61	VQRVVGSAAPASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVISVNMQTKSGGTR	120
QY	344	TGNVTLAEGPPA	355
Db	121	TGNVTLAEGPPA	132

Search completed: June 30, 2004, 16:55:18
Job time : 11.7991 secs

Qy 121 YDRTQDVAVLQLRGAGLPSAAITGGGVAVGEPVVMNGSGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLQLRGAGLPSAAITGGGVAVGEPVVMNGSGGGTTPRAVGRVVALGQTV 180
Qy 181 QASDSLTAETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFOLSGGGGPA 240
Db 181 QASDSLTAETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFOLSGGGGPA 240
Qy 241 IPIGQAWIAGQIRSGGSPTHIGPTAFGLGVVDNNGGARVQVVGSAAPASLGIST 300
Db 241 IPIGQAWIAGQIRSGGSPTHIGPTAFGLGVVDNNGGARVQVVGSAAPASLGIST 300
Qy 301 GDVITAVDGPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

US-10-193-002-80

; Sequence 80, Application US/10193002

; Publication No. US20030135026A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.

; Dillon, Davin C.

; Campos-Neto, Antonia

; Houghton, Raymond

; Vedvick, Thomas S.

; Twardzik, Daniel R.

; Lodes, Michael J.

; Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

; TUBERCULOSIS

; NUMBER OF SEQUENCES: 350

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/193,002

; FILING DATE: 10-Jul-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,596

; FILING DATE: 05-MAY-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.417C9

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 80:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 355 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 80:

US-10-193-002-80

Query Match

Best Local Similarity 100.0%; Score 1802; DB 14; Length 355;

Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSNRRRLRWSMLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
Db 1 MSNRRRLRWSMLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
Qy 61 PQVNVNITKLYNNNAVAGAGTIVIDPENGVLVTNNHVIAGATDINAFSVSGGQTVGVDDVVG 120
Db 61 PQVNVNITKLYNNNAVAGAGTIVIDPENGVLVTNNHVIAGATDINAFSVSGGQTVGVDDVVG 120
Qy 121 YDRTQDVAVLQLRGAGLPSAAITGGGVAVGEPVVMNGSGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLQLRGAGLPSAAITGGGVAVGEPVVMNGSGGGTTPRAVGRVVALGQTV 180
Qy 181 QASDSLTAETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFOLSGGGGPA 240
Db 181 QASDSLTAETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFOLSGGGGPA 240
Qy 241 IPIGQAWIAGQIRSGGSPTHIGPTAFGLGVVDNNGGARVQVVGSAAPASLGIST 300
Db 241 IPIGQAWIAGQIRSGGSPTHIGPTAFGLGVVDNNGGARVQVVGSAAPASLGIST 300
Qy 301 GDVITAVDGPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

US-10-084-843-79

; Sequence 79, Application US/10084843

; Publication No. US20030143243A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.

; Dillon, Davin C.

; Campos-Neto, Antonio

; Houghton, Raymond

; Vedvick, Thomas S.

; Twardzik, Daniel R.

; Lodes, Michael J.

; Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

; AND DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 355

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/084,843

; FILING DATE: 25-Feb-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,967

; FILING DATE: 05-MAY-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.411C9

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 79:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 355 amino acids

; TYPE: amino acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-084-843-79

Query Match      100.0%; Score 1802; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWWSLLSVAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWWSLLSVAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
QY 61 PQVWNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVDDVVG 120
DB 61 PQVWNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVDDVVG 120
QY 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCGAETTLNGLIOFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSGGGQGA 240
DB 181 QASDSLTCGAETTLNGLIOFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSGGGQGA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAPAAASLGIST 300
DB 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAPAAASLGIST 300
QY 301 GDVITAVDGAPINSATADALNGHHGPDVLSVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAPINSATADALNGHHGPDVLSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4
US-10-098-732A-2
; Sequence 2, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-C120100S
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: Mtb32A (Ra35FL)
US-10-098-732A-2

Query Match      100.0%; Score 1802; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWWSLLSVAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWWSLLSVAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
QY 61 PQVWNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVDDVVG 120
DB 61 PQVWNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVDDVVG 120
QY 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180

RESULT 5
US-09-712-363-161
; Sequence 161, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: Mtb363-161
US-09-712-363-161

Query Match      99.4%; Score 1792; DB 9; Length 355;
Best Local Similarity 99.4%; Pred. No. 9.6e-134;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWWSLLSVAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWWSLLSVAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
QY 61 PQVWNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVDDVVG 120
DB 61 PQVWNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVDDVVG 120
QY 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCGAETTLNGLIOFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSGGGQGA 240

```

Db 181 QASDLTGABETLNLQFDAAIQDGSQGVVNGLVGVVMTAASDNFQLSQGGGPA 240
QY 241 IPTQAMAIAGQIRSGGSGTTHIGTAFGLGVVDNNGNGARVQRVWGSAPAAISLGIST 300
Db 241 IPTQAMAIAGQIRSGGSGTTHIGTAFGLGVVDNNGNGARVQRVWGSAPAAISLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWOTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWOTKSGGTRTGNVTLAEGPPA 355
RESULT 6
US-09-886-349A-4
; Sequence 4, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-0090703
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35 mature)
US-09-886-349A-4

Query Match 90.7%; Score 1634; DB 12; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNVNINIKLGYNNVAGAGTGIVIDPNGVLT 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNVNINIKLGYNNVAGAGTGIVIDPNGVLT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVVGVDYDRTQDVAVLQLRGAGLPSAAITGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVVGVDYDRTQDVAVLQLRGAGLPSAAITGGVAVGEP 127
QY 153 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQDGSQGPV 212
Db 128 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQDGSQGPV 187
QY 213 VNLGVGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSGTTHIGTAFGL 272
Db 188 VNLGVGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSGTTHIGTAFGL 247
QY 273 GVVDNNGNGARVQRVWGSAPAAISLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332
Db 248 GVVDNNGNGARVQRVWGSAPAAISLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
QY 333 VNWOTKSGGTRTGNVTLAEGPPA 355
Db 308 VTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 7
US-10-098-732A-4
; Sequence 4, Application US/10098732A
; Publication No. US2003017529A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir

; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35 mature
US-10-098-732A-4
Query Match 90.7%; Score 1634; DB 14; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNVNINIKLGYNNVAGAGTGIVIDPNGVLT 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNVNINIKLGYNNVAGAGTGIVIDPNGVLT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVVGVDYDRTQDVAVLQLRGAGLPSAAITGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVVGVDYDRTQDVAVLQLRGAGLPSAAITGGVAVGEP 127
QY 153 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQDGSQGPV 212
Db 128 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQDGSQGPV 187
QY 213 VNLGVGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSGTTHIGTAFGL 272
Db 188 VNLGVGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSGTTHIGTAFGL 247
QY 273 GVVDNNGNGARVQRVWGSAPAAISLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332
Db 248 GVVDNNGNGARVQRVWGSAPAAISLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
QY 333 VNWOTKSGGTRTGNVTLAEGPPA 355
Db 308 VTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 8
US-10-369-983-19
; Sequence 19, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: wild-type
; OTHER INFORMATION: mature MTB32A (Ra35)

US-10-369-983-19

Query Match 90.7%; Score 1634; DB 15; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNNINIKLYNNVAGAGTGVVIDPENGVLIT 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNNINIKLYNNVAGAGTGVVIDPENGVLIT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSVGSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 127
QY 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
QY 213 VNGLGQVVMNTAASDNFQSGGGGFAIPIGQMAIAGQIRSGGSPVTHIGPTAFGL 272
DB 188 VNGLGQVVMNTAASDNFQSGGGGFAIPIGQMAIAGQIRSGGSPVTHIGPTAFGL 247
QY 273 GVVDDNNGNGARVQVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDIS 332
DB 248 GVVDDNNGNGARVQVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDIS 307
QY 333 VNWQTKSGGTRTGNVTTLAEGPPA 355
DB 308 VTWQTKSGGTRTGNVTTLAEGPPA 330

RESULT 9

US-09-886-349A-6

; Sequence 6, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCES: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA

Query Match 90.5%; Score 1631; DB 12; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNNINIKLYNNVAGAGTGVVIDPENGVLIT 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNNINIKLYNNVAGAGTGVVIDPENGVLIT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSVGSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 127
QY 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187

QY 213 VNGLGQVVMNTAASDNFQSGGGGFAIPIGQMAIAGQIRSGGSPVTHIGPTAFGL 272
DB 188 VNGLGQVVMNTAASDNFQSGGGGFAIPIGQMAIAGQIRSGGSPVTHIGPTAFGL 247
QY 273 GVVDDNNGNGARVQVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDIS 332
DB 248 GVVDDNNGNGARVQVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDIS 307
QY 333 VNWQTKSGGTRTGNVTTLAEGPPA 355
DB 308 VTWQTKSGGTRTGNVTTLAEGPPA 330

RESULT 10

US-10-098-732A-6

; Sequence 6, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA

Query Match 90.5%; Score 1631; DB 14; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNNINIKLYNNVAGAGTGVVIDPENGVLIT 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNNINIKLYNNVAGAGTGVVIDPENGVLIT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSVGSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEP 127
QY 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
QY 213 VNGLGQVVMNTAASDNFQSGGGGFAIPIGQMAIAGQIRSGGSPVTHIGPTAFGL 272
DB 188 VNGLGQVVMNTAASDNFQSGGGGFAIPIGQMAIAGQIRSGGSPVTHIGPTAFGL 247
QY 273 GVVDDNNGNGARVQVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDIS 332
DB 248 GVVDDNNGNGARVQVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHHPGDIS 307
QY 333 VNWQTKSGGTRTGNVTTLAEGPPA 355
DB 308 VTWQTKSGGTRTGNVTTLAEGPPA 330

RESULT 11

US-10-369-983-20

; Sequence 20, Application US/10369983
; Publication No. US20030235593A1

GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
PRIOR FILING DATE: 2003-02-18
PRIOR FILING DATE: 2003-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 20
LENGTH: 330
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:mutated
OTHER INFORMATION: MTB32MutSA (Ra35 mutSA)
US-10-369-983-20

Query Match 90.5%; Score 1631; DB 15; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVVIDPNGVVL 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVVIDPNGVVL 67
Qy 93 NNHVIAGATDINAFSVGSGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVGSGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 127
Qy 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPDGSGGPV 212
Db 128 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPDGSGGPV 187
Qy 213 VNLGQVVGWNTAAADNFQSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFI 272
Db 188 VNLGQVVGWNTAAADNFQSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFI 247
Qy 273 GVVDNNGGARVORVVGSAAPASLIGISTGDTVAVDGAIPINSATAMADALNGHHHPGDVIS 332
Db 248 GVVDNNGGARVORVVGSAAPASLIGISTGDTVAVDGAIPINSATAMADALNGHHHPGDVIS 307
Qy 333 VNWOTKSGGTRTGNVTLAEGPPA 355
Db 308 VTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 12
US-10-369-983-2
Sequence 2, Application US/10369983
Publication No. US20030235593A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
PRIOR FILING DATE: 2003-02-18
PRIOR FILING DATE: 2003-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 723
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:mutated
OTHER INFORMATION: MTB32MutSA (Ra35 mutSA)
US-10-369-983-20

OTHER INFORMATION: Description of Artificial Sequence:mutated
OTHER INFORMATION: MTB32-MTB39F fusion protein (MTB32MutSA)
US-10-369-983-2
Query Match 90.5%; Score 1631; DB 15; Length 723;
Best Local Similarity 99.1%; Pred. No. 1.3e-120;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVVIDPNGVVL 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVVIDPNGVVL 67
Qy 93 NNHVIAGATDINAFSVGSGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVGSGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 127
Qy 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPDGSGGPV 212
Db 128 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPDGSGGPV 187
Qy 213 VNLGQVVGWNTAAADNFQSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFI 272
Db 188 VNLGQVVGWNTAAADNFQSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFI 247
Qy 273 GVVDNNGGARVORVVGSAAPASLIGISTGDTVAVDGAIPINSATAMADALNGHHHPGDVIS 332
Db 248 GVVDNNGGARVORVVGSAAPASLIGISTGDTVAVDGAIPINSATAMADALNGHHHPGDVIS 307
Qy 333 VNWOTKSGGTRTGNVTLAEGPPA 355
Db 308 VTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 13
US-10-369-983-4
Sequence 4, Application US/10369983
Publication No. US20030235593A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
CURRENT FILING DATE: 2003-02-18
PRIOR FILING DATE: 2003-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 1010
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:MTB-102F fusion
OTHER INFORMATION: protein
US-10-369-983-4

Query Match 90.5%; Score 1631; DB 15; Length 1010;
Best Local Similarity 99.1%; Pred. No. 2e-120;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVVIDPNGVVL 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVVIDPNGVVL 67
Qy 93 NNHVIAGATDINAFSVGSGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVGSGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 127
Qy 153 VVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPDGSGGPV 212

Db 128 VVAMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPDAGGPV 187
QY 213 VNLGLGVGVMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 272
Db 188 VNLGLGVGVMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 247
QY 273 GVVNNNGGARVQVGVVGSAPAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 332
Db 248 GVVNNNGGARVQVGVVGSAPAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 307
QY 333 VNWQKSGGTRTGNVTNLAEGPPA 355
Db 308 VNWQKSGGTRTGNVTNLAEGPPA 330

RESULT 14

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 55.5%; Score 999.5; DS 15; Length 1016;
Best Local Similarity 84.1%; Pred. No. 26-70; 18; Indels 15; Gaps 4;
Matches 207; Conservative 6; Mismatches 18; Indels 15; Gaps 4;
QY 28 APAQAAPPALSQDRFADFPALDPSAMVAQVAPQVNNINTKLGYNNAVAGAGTGIVIDPN 87
Db 530 AAGDIAPPALSQDRFADFPALDPSAMVAQVAPQVNNINTKLGYNNAVAGAGTGIVIDPN 589
QY 88 GVLTNNHVIAGATDINAFSVSGGQTYGVYVVDRTQDVAVLQIRGAGGLPSAAIGGV 147
Db 590 GVLTNNHVIAGATDINAFSVSGGQTYGVYVVDRTQDVAVLQIRGAGGLPSAAIGGV 649
QY 148 AVGEPVAMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPD 207
Db 650 AVGEPVAMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPD 709
QY 208 SGGPVVNGLGQVVGMMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPT 267
Db 710 SGGPVVNGLGQVVGMMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPT 769
QY 259 SPTVHI 264
Db 764 SPAYVL 769

RESULT 15

US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-MTCC#2)
US-10-369-983-16

Query Match 55.0%; Score 992; DS 15; Length 1154;
Best Local Similarity 63.8%; Pred. No. 9.2e-70;
Matches 219; Conservative 17; Mismatches 59; Indels 48; Gaps 6;
QY 28 APAQAAPPALSQDRFADFPALDPSAMVAQVAPQVNNINTKLGYNNAVAGAGTGIVIDPN 87
Db 530 AAGDIAPPALSQDRFADFPALDPSAMVAQVAPQVNNINTKLGYNNAVAGAGTGIVIDPN 589
QY 88 GVLTNNHVIAGATDINAFSVSGGQTYGVYVVDRTQDVAVLQIRGAGGLPSAAIGGV 147
Db 590 GVLTNNHVIAGATDINAFSVSGGQTYGVYVVDRTQDVAVLQIRGAGGLPSAAIGGV 649
QY 148 AVGEPVAMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPD 207
Db 650 AVGEPVAMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPD 709
QY 208 SGGPVVNGLGQVVGMMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPT 267
Db 710 SGGPVVNGLGQVVGMMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPT 760
QY 268 AFLGLGVVDNNGGARVQVGVVGSAPAASLGISTGDIVITAVDGPINSATAMADALNGH 325
Db 761 ANDGV-----AAELTSAASVSVSVVSTLIVEPMMGPPAAAAAATAATPY 803
QY 326 HPGDVISVNW-----QTKSGGTRTGNVTNLAEGPPA 355
Db 804 -----VGWLAATAALAKETATQAAAAAFAFGTAFAMTVPPS 839

Search completed: June 30, 2004, 17:14:43
Job time : 28.6821 secs

1. The first part of the document is a list of the names of the persons who have been appointed to the various positions of the Board of Directors of the company. The names are listed in alphabetical order, and each name is followed by the position to which he or she has been appointed. The list is as follows:

2. The second part of the document is a list of the names of the persons who have been appointed to the various positions of the Board of Directors of the company. The names are listed in alphabetical order, and each name is followed by the position to which he or she has been appointed. The list is as follows:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 9.27806 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796c-4
Perfect score: 1802
Sequence: 1 MSNSRRSLRWKLLSVLAA.....QTKSGGTRTGNVTLAEPPA 355
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.78:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	99.4	355	2 F70983	probable serine pr
2	1306	72.5	361	2 S47170	hypothetical prote
3	1272.5	70.6	354	2 A87242	probable secreted
4	468.5	26.0	464	2 C70821	probable serine pr
5	460	25.5	382	2 H86930	probable secreted
6	460	25.5	452	2 T45448	probable serine pr
7	382	21.2	394	2 S74643	proteinase hhoA (E
8	372	20.6	407	2 A82150	serine proteinase
9	355.5	19.7	452	2 S77538	serine proteinase
10	354	19.6	433	2 H97199	htrA-like serine p
11	345	19.1	362	2 T35287	probable secreted
12	343	19.0	457	2 AG0433	proteinase (EC 3.4
13	341.5	19.0	514	2 A82591	periplasmic protei
14	335.5	18.6	429	2 A81894	serine proteinase
15	331	18.4	416	2 A82057	serine proteinase
16	331	18.4	441	2 E75357	probable periplasm
17	331	18.4	455	2 C91142	serine endoprotein
18	331	18.4	455	2 F85987	serine endoprotein
19	329.5	18.3	474	2 F83550	serine proteinase
20	329	18.3	398	2 B71284	probable periplasm
21	329	18.3	455	2 A80909	serine protease (E
22	328	18.2	355	1 J06082	trypsin-like prote
23	328	18.2	355	2 D91142	trypsin-like prote
24	327	18.1	352	2 E82307	trypsin-like prote
25	327	18.1	355	2 G85987	trypsin-like prote
26	326	18.1	455	2 JC6051	trypsin-like prote
27	322	17.9	475	1 S15337	heat shock protein
28	319	17.7	475	2 AC0528	protease DO precur
29	317.5	17.6	474	2 S45229	proteinase DO (EC

ALIGNMENTS

RESULT 1

F70983

probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002
C:Accession: F70983
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70983
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-355 <COL>
A:Cross-references: GB:Z95071; GB:AL123456; NID:G3242254; PIDN:CA809453.1; PID:G2181967
A:Experimental source: strain H37RV
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryps

Query Match	99.4%	Score 1792;	DB 2;	Length 355;
Best Local Similarity	99.4%;	Pred. No. 7.5e-98;		
Matches 353;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	MSNSRRSLRWKLLSVLAAVGLGLATAPAAAPALSDRFPADFPALPLDPSAMVAQVA	60	
Db	1	MSNSRRSLRWKLLSVLAAVGLGLATAPAAAPALSDRFPADFPALPLDPSAMVAQVG	60	
QY	61	PQVNVNTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSGSGGTGYGVDVVG	120	
Db	61	PQVNVNTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSGSGGTGYGVDVVG	120	
QY	121	YRQTQDVAVLQRGAGLPSAAIGGVAVGEFVAMGNSGGGGTTPRAVFGRWALGQTV	180	
Db	121	YRQTQDVAVLQRGAGLPSAAIGGVAVGEFVAMGNSGGGGTTPRAVFGRWALGQTV	180	
QY	181	QASDSLTGAEETLNGLIQFDAAIQDSCGPPVNGLGQVGVNNTAASDNFQLSQGGGFA	240	
Db	181	QASDSLTGAEETLNGLIQFDAAIQDSCGPPVNGLGQVGVNNTAASDNFQLSQGGGFA	240	
QY	241	IPFGQMAIAGQIRSGGSGPTVHIGFTAFGLGVVDNNGNGARVORVVGSAFAASLGIST	300	
Db	241	IPFGQMAIAGQIRSGGSGPTVHIGFTAFGLGVVDNNGNGARVORVVGSAFAASLGIST	300	
QY	301	GDVITAVDCAPINSATAMADALNGHHHPGDIVISNNWQTKSGGTRTGNVTLAEPPA	355	
Db	301	GDVITAVDCAPINSATAMADALNGHHHPGDIVISNNWQTKSGGTRTGNVTLAEPPA	355	

proteinase DO (EC
proteinase DO (EC
serine protease Do
htrA-like protein
proteinase DO (EC
serine proteinase
trypsin-like prote
proteinase hhoB (E
serine proteinase
global stress regu
serine proteinase
proteinase DO (EC
serine proteinase
probable periplasm

Db 355 AIPVDQAKRIADELSTGKA-----SHASLGQVQVINDKDTLGAKEIVVAGGAANAGV 408

QY 299 STGDVITAVDGPATNATAMADALNGHHGPDVIVSNWQTKSGGTRTGNVTLLAE 351

Db 409 PKGVVTKVDDREINSGADALVAARSKAPGATVALTFQDPGSGSRVQVTLGK 461

RESULT 5

H86930

Probable secreted serine proteinase [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002

C/Accession: H86930

R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holt, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, R.

A>Title: Massive gene decay in the leprosy bacillus.

A/Reference number: A86909; PMID:21128732; PMID:11234002

A/Accession: H86930

A>Status: preliminary

A:Molecule type: DNA

A/Residues: 1-382 <STO>

A/Cross-references: GB:AL450380; NID:gl3092536; PIDN:CAC29684.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML0176

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 25.5%; Score 460; DB 2; Length 382;

Best Local Similarity 36.4%; Pred. No. 5.3e-20;

Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGLATAPAAQAPALPSQDRFADFPALPLDPSAMVAQVA-----PQVNNINTKLGYNNNAV 77

Db 50 GAGPVTGPAASVPAANM-----PSGSVEQVAVKVPVSMVLETDLGRQSE-- 94

QY 78 AGTGIVDPNGVLTNNHVIAGA-----TDINAFSVGSQGYVGVVGYDR 123

Db 95 EGSVILSADGLIITNNHVAARSKAPGPGGGLSPKTTVTFF--DRTASTFVVGADP 151

QY 124 TDVAVLQRLGAGGLPSAAIGGG--VAVGEPVWAMGNSGGGTPPRAVPGRVVAGLQTVQ 181

Db 152 TSDIAVVRVQISGLTPTMGSSADLRVGPVAVGSPGLAGT--VTSGIVSALNRPVS 209

QY 182 ASDSLTGAEETLNGLIQFDAAIQDGSQGPVNGLGQVVGNTAA-----SDNFQLSQG 235

Db 210 TTGE-SGNQNTVLDAIQTDAAINPNSGGALVNMGGQLVGNVSAIATLGADSGDAQSGSI 268

QY 236 GQGAIPITGOAMATAGIQRSGGSPVTHIGPTAFGLGVDVNNNGN-GARVQVVGSPAPAA 294

Db 269 GLGFAIPVDQAKRIADELSTG--KATH-----ASLGQVQVATDKTPGAKVMDVVAGGAAA 322

QY 295 SLGISTGDVITAVDGPATNATAMADALNGHHGPDVIVSNWQTKSGGTRTGNVTLLAE 351

Db 323 NAAVPKGVVLTQVDDRLISSADALVAARSKAPGDKVSLTYQDQSGSRVQVTLGK 379

RESULT 6

T45448

Probable serine proteinase (EC 3.4.21.-) MLCB373.28 [similarity] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Dec-2002

C/Accession: T45448

R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, February 1999

A/Reference number: Z22967

A/Accession: T45448

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A/Residues: 1-452 <JAM>

A/Cross-references: EMBL:AL035500; PIDN:CAB36690.1

A/Experimental source: cosmid L373

Query Match 21.2%; Score 382; DB 2; Length 394;

Best Local Similarity 31.8%; Pred. No. 2e-15;

Matches 116; Conservative 50; Mismatches 117; Indels 84; Gaps 14;

QY 13 WLLSLVAALVGLGLATAP-----AQAAPPA-----LSQDRFADFPALPLDPSAMVA 57

Db 13 YLLAFVAGTAFGTANLPHAVAAADDLPAPVITAQAQSVLTSSEFV-----AAAAS 63

QY 58 QVAPQVNNITK-----LQYNNAV-----GAGTGVIDPENGVV 90

Db 64 RSGFAVVRIDTETVWTRTDPILDDPFFOEFGRSPFPVPRRIAGQSGGFIDNSGII 123

QY 91 LTNNHVIAGTADINAFSVGSQGYVGVVGYDRYQDVAVLQRLGAG-GLPFAAIG--GGV 147

C:Genetics:

A>Note: MLCB373.28

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

C:Keywords: hydrolase; serine proteinase

F:182,224,305/Active site: His, Asp, Ser #status Predicted

Query Match 25.5%; Score 460; DB 2; Length 452;

Best Local Similarity 36.4%; Pred. No. 6.3e-20;

Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGLATAPAAQAPALPSQDRFADFPALPLDPSAMVAQVA-----PQVNNINTKLGYNNNAV 77

Db 120 GAGPVTGPAASVPAANM-----PSGSVEQVAVKVPVSMVLETDLGRQSE-- 164

QY 78 AGTGIVDPNGVLTNNHVIAGA-----TDINAFSVGSQGYVGVVGYDR 123

Db 165 EGSVILSADGLIITNNHVAARSKAPGPGGGLSPKTTVTFF--DRTASTFVVGADP 221

QY 124 TDVAVLQRLGAGGLPSAAIGGG--VAVGEPVWAMGNSGGGTPPRAVPGRVVAGLQTVQ 181

Db 222 TSDIAVVRVQISGLTPTMGSSADLRVGPVAVGSPGLAGT--VTSGIVSALNRPVS 279

QY 182 ASDSLTGAEETLNGLIQFDAAIQDGSQGPVNGLGQVVGNTAA-----SDNFQLSQG 235

Db 280 TTGE-SGNQNTVLDAIQTDAAINPNSGGALVNMGGQLVGNVSAIATLGADSGDAQSGSI 338

QY 236 GQGAIPITGOAMATAGIQRSGGSPVTHIGPTAFGLGVDVNNNGN-GARVQVVGSPAPAA 294

Db 339 GLGFAIPVDQAKRIADELSTG--KATH-----ASLGQVQVATDKTPGAKVMDVVAGGAAA 392

QY 295 SLGISTGDVITAVDGPATNATAMADALNGHHGPDVIVSNWQTKSGGTRTGNVTLLAE 351

Db 393 NAAVPKGVVLTQVDDRLISSADALVAARSKAPGDKVSLTYQDQSGSRVQVTLGK 449

RESULT 7

S74643

proteinase hhoA (EC 3.4.-.-) - Synechocystis sp. (strain PCC 6803)

N:Alternate names: protein sll1679

C:Species: Synechocystis sp.

A/Variety: PCC 6803

C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002

C/Accession: S74643

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996

A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis s.

A/Reference number: S74322; PMID:97061201; PMID:8905231

A/Accession: S74643

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A/Residues: 1-394 <KAN>

A/Cross-references: EMBL:D90900; GB:AB001339; NID:gl651768; PIDN:BAAL6795.1; PID:gl65186f

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: hhoA

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

C:Keywords: hydrolase; proteinase

Query Match 21.2%; Score 382; DB 2; Length 394;

Best Local Similarity 31.8%; Pred. No. 2e-15;

Matches 116; Conservative 50; Mismatches 117; Indels 84; Gaps 14;

QY 13 WLLSLVAALVGLGLATAP-----AQAAPPA-----LSQDRFADFPALPLDPSAMVA 57

Db 13 YLLAFVAGTAFGTANLPHAVAAADDLPAPVITAQAQSVLTSSEFV-----AAAAS 63

QY 58 QVAPQVNNITK-----LQYNNAV-----GAGTGVIDPENGVV 90

Db 64 RSGFAVVRIDTETVWTRTDPILDDPFFOEFGRSPFPVPRRIAGQSGGFIDNSGII 123

QY 91 LTNNHVIAGTADINAFSVGSQGYVGVVGYDRYQDVAVLQRLGAG-GLPFAAIG--GGV 147

```

Db      124 LTNHVVVDGASKV-VVTLRDGRFTDGVGTDEVTDLAVVKIEPQGSALFPVAPLGTSSNL 182
QY      148 AVGEPPVVMGNSGGOGTTPRAVGRVVALG--QTVOASDSLTCABETLNGLLQFPAATOP 205
Db      183 QVGDWIAVGNPVLGNT-----VTLGISTLGRSAAQAGIPDKRVEFTQTDAAINP 234
QY      206 GDSGGPVVNLGQVGVGNTAASDNFOLSGOGGFAIPICQAMAIAGQIRSGSGSPVTHIG 265
Db      235 GNSGGPLLNARGVEIGINTA---IRADATGIGFAIPIDQAKAIQNTLAAAGTVPHPYIG 290
QY      266 PTAFGLGLGV---VDNNGN-----GARVQRVVGSAAPASLGISTGVDVITAVDGAP 311
Db      291 -VOMNITVDQAQONNRNPNPFIPEVDGILVMRVLPCTPAERAGIRRGDVIVAVDGT 349
QY      312 INSATAM 318
Db      350 ISDGAEL 356

RESULT 8
AG2150
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2150
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:213595285; PMID:11759840
A:Accession: AG2150
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074457.1; PID:g17131851; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: dir2758
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryg

```

```

Query Match      20.6%; Score 372; DB 2; Length 407;
Best Local Similarity 30.3%; Pred. No. 7,8e-15;
Matches 125; Conservative 59; Mismatches 141; Indels 88; Gaps 15;

QY      7 RSLR--WSWLLSVLAAGL---GLATAPQAAP-----PALSDRFPADFPALPLD 51
Db      10 RSIQLGTHLAIPTGIVVLTVSSLRVLPQAEAPAPNPSTTGSAPELVAQRQSPTAAVIGN 69
QY      52 PS---AMVAQVAPVNNITKLGNNAV-----GAGTG 81
Db      70 SSFVTAANVRVGSVAVRDTERTITRRVDPLEDPFRFRPFGEFGQQLPEQVRGLGSG 129
QY      82 IVIDPNGVLLNNHVIAGATDINAFSGSGQTYGVVDVDRDQDVAVLQRLGAGGLPSA 141
Db      130 FIIDKSGILTNHVVDRKADRTV-RLKDGSRFGKQGVDEVTDLAVKINAGNSLFA 188
QY      142 AIG--GGVAVGEPPVVMGNSGGOGTTPRAVGRVVALG--QTVOASDSLTCABETLNGLI 197
Db      189 PLGSSNNVQVGDWIAVGNPLGFNT-----VTLGIVSTLKRSAQGVITDKRLDFI 240
QY      198 QFDAAIQPGDSGGPVNGLGVGVMGNTAASDNFOLSGOGGFAIPICQAMAIAGQIRSGG 257
Db      241 QTDAAINPGNSGGPILLNDKGEVIGINTA---IRADAWGIGFAIPIDKAKAIATQLERDG 296
QY      258 GSPVTHIGPTAFILGV-----VDNNG-----NGARVQRVVGSAAPASLGI 298
Db      297 --KVH-----PYLGVQWATITPELAQQNNIDPNPSAFAIPEVNGVLRVVPNSPFAANAGI 350
QY      299 STGDVITAVDGAIPINSATAMADALNGHHGPDVIVSNWQTKSGGTRTGNVTLAE 351
Db      351 RRGDVILLQVGDQALTTAEQLQNVVNSRLQALQVRLQ-RGNQTQQLSVRTAE 402

```

```

RESULT 9
S77538
serine proteinase (EC 3.4.21.-) htra - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein slr1204
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002
C:Accession: S77538
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S77538
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-452 <KAN>
A:Cross-references: EMBL:D90905; GB:AB001339; NID:g1652360; PIDN:BA017385.1; PID:g165246;
A:Note: the nucleotide sequence was submitted to the EMBL data library, June 1996
C:Genetics:
A:Gene: htra
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps
C:Keywords: hydrolase; serine proteinase

Query Match      19.7%; Score 355.5; DB 2; Length 452;
Best Local Similarity 29.2%; Pred. No. 8.1e-14;
Matches 116; Conservative 65; Mismatches 129; Indels 87; Gaps 16;

QY      16 SVLAAVGLGLATAPQAAPALSDRFPADFPALPLDPSAMVAQV----APQVYNNITKLIG 71
Db      84 SVISPLVTNQSIAPANESLATNLQSLRS-----PREPSNFVVDVVESTGPAVRINAOKT 138
QY      72 YNNAV-----GAGTGIVDPNGVVLNNHVIAGATDIN 104
Db      139 VKSQVPEAFNDPFLRFFGSGQMPNERNVQRGTSGFVSDGKIFTNARHVVDGADEVT 198
QY      105 AFSVSGSGQTYGVVDVGVYDRTQDVAVLQRLGAGGLPSAATGG--VAVGEPPVVMGNSGGQ 162
Db      199 V-TLKDGSRFPGRVNGSDPSTDVAVKIE-AGDLFTVALGSDHLOVGEWAIAGNPJGL 256
QY      163 GGTTPRAVGRVVALGQTVQASDSLTCABETLNGLLQFPAATOPGDSGGPVNGLQGVVM 222
Db      257 DNT--VTTGILSATGR--RSAD--IGVPDKRVEFTQDAAINPGNSGGPFLNADQGVIGM 310
QY      223 NTAASDNFOLSGOGGFAIPICQAMAIAGQIRSGSGSPVTHIGPTAFILGLGVVDNN--- 278
Db      311 NTAIONAQ---GIGFAIPINKAEIAQOLIATG--KVEH-----AYLGIQWVTWTPBLQ 360
QY      279 -----GNGARVQRVVGSAAPASLGISTGVDVITAVDGAIPINSATAMADALNGH 325
Db      361 SOIROETGNINIPDKRGVIMQVMPNSPAAIAKLEQGVLSLQGVQVQVVENAQVQSLVGL 420
QY      326 HPGDVISVNWQTKSGTGTG---NVTLAG-----PP 354
Db      421 AVGDEVEL-----GILNGQQNLTVTIGALPSAPP 451

RESULT 10
H97199
htra-like serine protease (with PDZ domain) [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Dec-2002
C:Accession: H97199
R:Knoll, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97199
A:Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-433 <RUR>
A:Cross-references: GB:AE001437; PIDN:AAK0387.1; PID:G15025449; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics: CAC2433
C:Superfamily: Bacterichia coli trypsin-like proteinase degS; GLGF domain homology; tryP

Query Match 19.6%; Score 354; DB 2; Length 433;
Best Local Similarity 32.0%; Pred. No. 9.4e-14;
Matches 101; Conservative 51; Mismatches 110; Indels 54; Gaps 10;

QY 50 LDFSMVAQVAPGVNINIKL-----GYNNAVAGAGTGIVDPNGVILT 92
DB 122 LTVSQIVKVPSPAVGVSTKTITQNDFFSGSSNGSSSTQEGMGSLIFNNDGYILT 181
QY 93 NNHVIAGATDINAFSGSGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIGGVA---V 149
DB 182 NYHVIRKADKL-AVILNNKEVSAKVNYDEANDIAVIRKVTGFTVPGVAELGSSASLV 240
QY 150 GEPVAVNGSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQEDAAIQDSDG 209
DB 241 GDSVVAIGNPLGREFLTGTVTVGVSAVNRVAVSE---GQKQT---YIQTDAINFGNSG 294
QY 210 GPVYNGIGGVVGMNTAASDFQISGGQ---GFAIPIGQAMALAGQIRSGGSGPTVHIGP 266
DB 295 GPLVNSFGQVVGNSA-----KISENGVEGIGHSIDITVKSQIQLSK-----P 339
QY 267 TAFILGLG--VVD-----NNGNGARVQVRVVGSAASLGISTGDIVITAVDGPAPNSATA 317
DB 340 ILMGISGEAVDKSTAEQHNIPQGVYIEIQDFSSAQKAGMQGVGDVITKFDGKKVSTSD 399
QY 318 MADALNGHHFGDVISV 333
DB 400 IDSISKXNSGDTVQV 415

RESULT 11
T35287
probable secreted proteinase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Dec-2002
C:Accession: T35287
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: 221574
A:Accession: T35287
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-362 <SBE>
A:Cross-references: EMBL:AL096872; PIDN:CAB51253.1; GSPDB:GN000070; SC0DB:SC5F7.30
A:Experimental source: strain A3(2)
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryP

Query Match 19.1%; Score 345; DB 2; Length 362;
Best Local Similarity 32.6%; Pred. No. 2.6e-13;
Matches 125; Conservative 49; Mismatches 154; Indels 56; Gaps 14;

QY 1 MSNRRRLRWSLLSVLAAGLGLATA-----PAQAPALSDQR 41
DB 1 MDTSRLLRLLPVAVPACV-LLIATGCSAGAGTDRGSGSAREGDTAQAAAPRAASEL 59
QY 42 PADFPALPLDPSAMVAQVAPVNTKLYGYNNAVAGTGIVDPNGVLTNNHVIAGAT 101
DB 60 EADY-----ERVIKDVLPSVVIQIA-----GDSLGSGVYDDKGVHVTNAHVVG--- 103
QY 102 DINAFSVSGSGQTYG---VDVYGYDRTQDVAVLQL-RGAGGLPSA--AIGGGVAVGEPVVA 155
DB 104 DAKSPRVTTARTEGALTAKLVSSYPEQDLAVIKLDKVPENRAARFADSAKVEVGQIVLA 163
QY 156 MNSGSGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQPDAAIQDSDGSPVYNG 215

DB 164 MGSPLGLSSS--VTQGISVATGRVITSGSGGTGATIANVQVTSAAIPGNSGALVNL 221
QY 216 LGQVVGWNTAASDNFQISQG---CGGPAIPIGQAMALAGQIRSG---GSPTVHIGTAP 269
DB 222 DGQVIGIPTLAATDPGLGDSAAPGIGFPAIPASWVTTVAGQIVRDGKVTDSGRAALGITA- 280
QY 270 LGLGVNDNGN--GARVQVRVGSAPASLGISTGDIVITAVDGPAPNSATAMADALNGHHP 327
DB 281 --RTVDDSYRPAAGAAVVEVSDGGAADAGLRPGLVGLGDTDTITTSISEALASMRP 338
QY 328 GDVISVNMQTKSGGTRTGNVTLAE 351
DB 339 GDRTKVTV-TRDGKEHTAEVTLGE 361

RESULT 12
AG0433
proteinase (EC 3.4.21.-) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG0433
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0433
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-457 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92795.1; PID:G15981488; GSPDB:GN00175
C:Genetics:
C:Gene: degQ
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 19.0%; Score 343; DB 2; Length 457;
Best Local Similarity 31.2%; Pred. No. 4.4e-13;
Matches 122; Conservative 53; Mismatches 126; Indels 90; Gaps 18;

QY 12 SWLLSVLA-AVCLGLATAP---AQAAPPALSDQRADFPALPLDPSAMVAQVAPVNI- 66
DB 5 SLLLSALATSVGLGLASPMVMSAALPAVAGQ---TLPSL-----APMLEKVLPAVSVH 57
QY 67 -----NTKLYGYNNA---VGAGTGIVID-PNGVVLTKNNHVIAGAT 101
DB 58 VSGSQAQQLPEEFKFFGPNAPSGKSSRRPFGLSGVINAEBGYILTNNHINNAD 117
QY 102 DINAFSVSGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAI--GGVAVGEPVAVGNS 159
DB 118 KIRV-QLNDGREYDAKLLGRDEQTDIALQLTDAKNLTAIKTADSDNRVGFAPAVAGNP 176
QY 160 GSGGTTPRVPGRVVALGQTVQASDSLTGAE-ETLNGLIQPDAAIQDSDGSPVYNGLQ 218
DB 177 FGLGQT--ATSGLIISALR-----SGLNLEGLNFQTDASINRNGSGGALVNLGE 226
QY 219 VGMNTAASDNFQISQGG---GFAIPIGQAMATA-----GQIRSG-----GSPTVH 263
DB 227 LIGINTAI-----LAPCGGNIGIGFPAIPSNMAQNLISQQLIEFGEVKRGLGIRGSEMTAD 281
QY 264 IGTAFLGLGVNDNGNGARVQVRVGSAPASLGISTGDIVITAVDGPAPNSATAMADALN 323
DB 282 IAKAFNI-----DAQRGAPVSEVLPKSAKAKGPKGVDVLISVDGKKISSFAELRAKVG 335
QY 324 GHHPGDIVISVNMQTKSGGTRTGNVTLAEPPP 354
DB 336 TTGFGTKIKIG-----LLREGKP 353

RESULT 13
A82581

periplasmic proteinase XP2241 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
R:Accession: AB2581
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: A82581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <SIM>
A:Cross-references: GB:AE004037; GB:AE003849; NID:g9107394; PIDN:AAF85040.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpon, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: Xf2241
C:Superfamily: Helicobacter serine proteinase

Query Match 19.0%; Score 341.5; DB 2; Length 514;
Best Local Similarity 31.6%; Pred. No. 6.1e-13;
Matches 123; Conservative 55; Mismatches 128; Indels 83; Gaps 17;
QY 3 NSRRSLRWMLLSV---LAAVGLGLATAPAAAP--PALSDQDFADFPALDEL-----DP 52
DB 2 NSRIHT-RCFGLLAITLPLAACGAOHNSPTTAPSTPIVTP---STTPAPQLVAGLPDF 57
QY 53 SAMVAQVAPVNVNT-----KLGYNNAV----- 76
DB 58 TQVDVQGVGVNIVETVTRKKVGRKGIPLDNIPIEFRRFFGDFQMPNQPRGGQDE 117
QY 77 -----GAGTGIVDPNGVLTNNHVIATGADTINAFSGQTYGVVDVDRTPQDVAVL 130
DB 118 GGIAGRGMSGFIISKGVILTNHHVITGASEV-TIKLTDREPKAKIIGSDQYDVALL 176
QY 131 QLRGAGGLPSAAGGVAV--GEPPVAMNSGGGGTTPRVPGRVVALGQTVQASDSLGT 188
DB 177 KI-DANKNLTPIVTRIGSSSLKGGVWVAIGSPFLDHSVTA--GIVSALGRS-----TS 226
QY 189 AEETLNGLIQDAATIQDSDGGPVVNGLGQVGVNMTAASNFQLSQG--GQGFAPITGQA 246
DB 227 DDQRVVPPIQTDVPINQNSGGPILLNTEGEVIGIN---SQIFSASGVMGIFSFAIPNL 283
QY 247 MAIAGQIRSGG-----GSPTVHIGTAFI---GLGVVNNNGGARVQVRVGSAPASLGIS 299
DB 284 INAAEQIRKTKGVQMSRLMGVEIGIPIDAKQAGLGLPDS--RGALVNNIPPHSPAAKAGIE 341
QY 300 TGDVITAVDGPINSATAMADALNGHHPG 328
DB 342 VGDVIRSVNGKVISFSFDLPPLICMMPGP 370

RESULT 14
AD1894
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD1894

Query Match 18.4%; Score 331; DB 2; Length 416;
Best Local Similarity 31.5%; Pred. No. 2e-12;
Matches 92; Conservative 57; Mismatches 115; Indels 28; Gaps 10;
QY 335 WQTKSGGTRTG-NVTILA 350
DB 407 W-----RNGRNLNLA 416
RESULT 15
AB2057
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2057
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073707.1; PID:g17131098; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2008
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypt

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1894
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-429 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073659.1; PID:g17130047; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2002
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypt

Query Match 18.6%; Score 335.5; DB 2; Length 429;
Best Local Similarity 29.2%; Pred. No. 1.1e-12;
Matches 110; Conservative 59; Mismatches 137; Indels 71; Gaps 14;
QY 18 LAAVGLGLATAPAAQAPALSDQDFADFPALPLDPSAMVAQVAPVNVNNTKLGYNNAV- 76
DB 67 LAQKASDLAVSRVDAAPP--LGNTDINF-----VTQVVRVGPVAVVREASRTVTSLRP 119
QY 77 -----GAGTGIVDPNGVLTNNHVIATGADTINAFSGVSG 111
DB 120 AEFNDPFFERFFGSLPQQQERVQRTGSGFLISADGSLTNAHVVDGADTVRVI-LKDG 178
QY 112 QTVGVVGVYDRTQDVAVQLRGAGGLPSAAITGGG--VAVGPPVAMNSGGQGTTPRAV 169
DB 179 RSFQGVLTGTDNLTDVAVVKIQ-ANNLPTLTGVNSDQLQFGQWAIAGNPLGLDNT--VT 235
QY 170 PGRVWALGOTVQASDSLTCAGETLNGLIQFDDAAIOPGSGGPPVNVNGLGQVGVNMTAASDN 229
DB 236 TGIISATGT-----SNQIGAPKRVETIOTDAINPNSGGPLIYRGEVIGNVTAIIQG 291
QY 230 FQLSQGGQGFAPIGQAMAIAGQIRSGGSPVHIGTAFGL-----GVV 275
DB 292 AQ-----GLGFAIPKTAQIRSNQLIATGKVPYLG-IQWVGLTPQIKQINSDNSGLT 346
QY 276 DNNNGCARVORVVGSAASLGISTGDTVITAVDGPINSATAMADALNGHHPGCDVISVN- 334
DB 347 VDRKGVLVVRVLPNSPAARAGLRAGDVITQKLVNGQAVTASNVQRAVENAQVGGQLLEL 406
QY 335 WQTKSGGTRTG-NVTILA 350
DB 407 W-----RNGRNLNLA 416

Query Match 18.6%; Score 335.5; DB 2; Length 429;
Best Local Similarity 29.2%; Pred. No. 1.1e-12;
Matches 110; Conservative 59; Mismatches 137; Indels 71; Gaps 14;
QY 18 LAAVGLGLATAPAAQAPALSDQDFADFPALPLDPSAMVAQVAPVNVNNTKLGYNNAV- 76
DB 67 LAQKASDLAVSRVDAAPP--LGNTDINF-----VTQVVRVGPVAVVREASRTVTSLRP 119
QY 77 -----GAGTGIVDPNGVLTNNHVIATGADTINAFSGVSG 111
DB 120 AEFNDPFFERFFGSLPQQQERVQRTGSGFLISADGSLTNAHVVDGADTVRVI-LKDG 178
QY 112 QTVGVVGVYDRTQDVAVQLRGAGGLPSAAITGGG--VAVGPPVAMNSGGQGTTPRAV 169
DB 179 RSFQGVLTGTDNLTDVAVVKIQ-ANNLPTLTGVNSDQLQFGQWAIAGNPLGLDNT--VT 235
QY 170 PGRVWALGOTVQASDSLTCAGETLNGLIQFDDAAIOPGSGGPPVNVNGLGQVGVNMTAASDN 229
DB 236 TGIISATGT-----SNQIGAPKRVETIOTDAINPNSGGPLIYRGEVIGNVTAIIQG 291
QY 230 FQLSQGGQGFAPIGQAMAIAGQIRSGGSPVHIGTAFGL-----GVV 275
DB 292 AQ-----GLGFAIPKTAQIRSNQLIATGKVPYLG-IQWVGLTPQIKQINSDNSGLT 346
QY 276 DNNNGCARVORVVGSAASLGISTGDTVITAVDGPINSATAMADALNGHHPGCDVISVN- 334
DB 347 VDRKGVLVVRVLPNSPAARAGLRAGDVITQKLVNGQAVTASNVQRAVENAQVGGQLLEL 406
QY 335 WQTKSGGTRTG-NVTILA 350
DB 407 W-----RNGRNLNLA 416

Query Match 18.4%; Score 331; DB 2; Length 416;
Best Local Similarity 31.5%; Pred. No. 2e-12;
Matches 92; Conservative 57; Mismatches 115; Indels 28; Gaps 10;
QY 335 WQTKSGGTRTG-NVTILA 350
DB 407 W-----RNGRNLNLA 416
RESULT 15
AB2057
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2057
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073707.1; PID:g17131098; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2008
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypt

Query Match 18.4%; Score 331; DB 2; Length 416;
Best Local Similarity 31.5%; Pred. No. 2e-12;
Matches 92; Conservative 57; Mismatches 115; Indels 28; Gaps 10;
QY 335 WQTKSGGTRTG-NVTILA 350
DB 407 W-----RNGRNLNLA 416
RESULT 15
AB2057
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2057
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073707.1; PID:g17131098; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2008
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypt

Query Match 18.4%; Score 331; DB 2; Length 416;
Best Local Similarity 31.5%; Pred. No. 2e-12;
Matches 92; Conservative 57; Mismatches 115; Indels 28; Gaps 10;
QY 335 WQTKSGGTRTG-NVTILA 350
DB 407 W-----RNGRNLNLA 416
RESULT 15
AB2057
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2057
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073707.1; PID:g17131098; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2008
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypt

Query Match 18.4%; Score 331; DB 2; Length 416;
Best Local Similarity 31.5%; Pred. No. 2e-12;
Matches 92; Conservative 57; Mismatches 115; Indels 28; Gaps 10;
QY 335 WQTKSGGTRTG-NVTILA 350
DB 407 W-----RNGRNLNLA 416
RESULT 15
AB2057
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2057
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073707.1; PID:g17131098; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2008
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypt

QY	77	GAGTGIIVDPNGVLTNNHVIAGTDINAFSVGSGQTYGVVGVYDRTQDVAVLQLRGAG	136
Db	133	SGSGFIISSSGQIITNAHVVDGDEVTV-TLKGRSPDGKVLGEDPVTDVAVIQI-NAN	190
QY	137	GLPSAAIGGGVAV--GEPVAMNSGGQGTTPRAYGRVVALGQTVQASDSLTAETLN	194
Db	191	NLPTVAVGNSEVLQFGEAVIAIGNPLGLNNS-VTSGIISATGR---SGSDIGASDKRV	244
QY	195	GLIQDAAIQDGGPGPVVNGLGQVVGWNTAASDNFOLSGGQGFAPIGQAMAIAGQIR	254
Db	245	DYLOTDAAINPGNSGGPLNARGQVIGMNTAIIQGAQ---GLGFAIPINTVQKVSQELI	300
QY	255	SGGSGPTVHIG-PTAFLGLGVVD--NNGNGARVQ-----RVVGSAPAASLGISTG	301
Db	301	TQGVDPHPYLGVMATLTPQVKERINEREGDRINITADRGVLLVRIVPGSPAANAGLRPG	360
QY	302	DVITAVDGAIPINSATAMADALNHHHPGDVISVNWOTKSGGTRTGNVTLAEGP	353
Db	361	DIITOSINNSQVTTVEQVKIVENSQIQPLQICIE-RNGQTTQVNVSPAPLP	411

Search completed: June 30, 2004, 16:53:57
Job time : 13.2781 secs

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The President talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

2. The second part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The President talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

3. The third part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The President talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

4. The fourth part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The President talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

5. The fifth part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The President talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 5.77978 Seconds
(without alignments)
3198.204 Million cell updates/sec

Title: US-09-597-796c-4
Perfect score: 1802
Sequence: 1 MSNSRRSLRWMLLSVLA.....QTKSGGTRTCNVTLAEGPPA 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	350.5	19.5	458	1	YVTA_BACSU Q9r911 bacillus su
2	328	18.2	355	1	DEGS_ECOLI P31137 escherichia
3	326	18.1	455	1	DEGQ_ECOLI P39099 escherichia
4	322	17.9	475	1	DEGP_SALTY P26982 salmonella
5	321	17.8	413	1	HTRA_LACHE Q24h7 lactobacill
6	317.5	17.6	474	1	DEGP_ECOLI P09376 escherichia
7	310	17.2	408	1	HTRA_LACIA Q91a06 lactococcus
8	306	17.0	513	1	DEGP_BRUME Q8vg32 brucella me
9	305	16.9	513	1	DEGP_BRUSU Q4597 brucella su
10	295.5	16.4	449	1	HTRA_BACSU Q34358 bacillus su
11	295	16.4	466	1	P45129 haemophilus
12	291	16.1	504	1	DEGP_RHIME Q52894 rhizobium m
13	290	16.1	503	1	DEGP_BARHE P54925 bartorella
14	289.5	16.1	437	1	DEG1_ARATH Q22609 arabidopsis
15	288	16.0	497	1	DEGP_CHLTR P18594 chlamydia t
16	286.5	15.9	497	1	DEGP_CHLMU Q9pl97 chlamydia m
17	286	15.9	478	1	DEGP_EUCAI P57322 buchnera ap
18	283	15.7	508	1	DEGP_RICCN Q92ja1 rickettsia
19	282	15.6	488	1	DEGP_CHLPN Q26t0 chlamydia p
20	276.5	15.3	400	1	YVXA_BACSU P39668 bacillus su
21	276	15.3	448	1	DEG8_ARATH Q91u10 arabidopsis
22	272.5	15.1	478	1	DEGP_EUCAP Q85291 buchnera ap
23	263	14.6	458	1	HRA2_HUMAN Q43464 homo sapien
24	258	14.3	513	1	DEGP_RICPR Q05942 rickettsia
25	254	14.1	458	1	HRA2_MOUSE Q9j1y5 mus musculus
26	244.5	13.6	340	1	DEGS_HAETN P44947 haemophilus
27	238.5	13.2	453	1	HRA3_HUMAN P83110 homo sapien
28	235	13.0	480	1	HRA1_HUMAN Q92743 homo sapien
29	231.5	12.8	476	1	HRA4_HUMAN P83105 homo sapien
30	225.5	12.5	460	1	Q9d236 mus musculus
31	225	12.5	480	1	HRA1_MOUSE Q9r118 mus musculus
32	178.5	9.9	321	1	SEPA_ARATH Q9se17 arabidopsis
33	173	9.6	630	1	Y4BJ_RHISN P55377 rhizobium s

ALIGNMENTS

RESULT 1

ID	YVTA_BACSU	STANDARD;	PRT;	458 AA.
AC	Q9r911; Q35021; Q35039;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Probable serine protease Yvta (EC 3.4.21.-).			
GN	YVTA OR YVTF OR BSU33000.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]_TaxID=1423;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=20158875; PubMed=10692364;			
RA	Noone D., Howell A., Devine K.M.;			
RT	"Expression of ykdA, encoding a Bacillus subtilis homologue of HtrA,			
RT	is heat shock inducible and negatively autoregulated.";			
RL	J. Bacteriol. 182:1592-1599 (2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98015415; PubMed=9353931;			
RA	Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;			
RT	"Sequencing of regions downstream of addA (98 degrees) and citG (289			
RT	degrees) in Bacillus subtilis.";			
RL	Microbiology 143:3305-3308 (1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogaawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Pujic P., Furnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiuchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,			
RA	Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			

34	162.5	9.0	457	1	PRTC_STRGR	P52320 streptomyce
35	158.5	8.8	747	1	ELS_BOVIN	P04985 bos taurus
36	153	8.5	801	1	Y747_MYCTU	O53810 mycobacteri
37	150.5	8.4	515	1	Y140_MYCTU	O50594 mycobacteri
38	147	8.2	299	1	PRTB_STRGR	P00777 streptomyce
39	146.5	8.1	864	1	ELS_RAT	Q09372 rattus norv
40	145	8.0	914	1	WA22_MYCTU	O06794 mycobacteri
41	142.5	7.9	957	1	Y278_MYCTU	P56877 mycobacteri
42	141.5	7.9	710	1	F1B1_ADEGL	Q64761 avian adeno
43	141	7.8	2021	1	OMPA_RICCN	Q52657 rickettsia
44	139.5	7.7	778	1	YQ34_MYCTU	E71933 mycobacteri
45	139.5	7.7	2364	1	PGCA_BOVIN	F13608 bos taurus

RA Viari A., Wambutt R., Medler E., Medler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 RN [4]
 RN TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=20576168; PubMed=11133960;
 RA Noone D., Howell A., Collier R., Devine K.M.;
 RT *YkdA* and *YvtA* HtrA-like serine proteases in *Bacillus subtilis*.
 RT engage in negative autoregulation and reciprocal cross-regulation of
 RT *ykdA* and *yvtA* gene expression.";
 RL J. Bacteriol. 183:654-663(2001).
 CC -!- FUNCTION: May be involved in processing, maturation, or secretion
 CC of extracellular enzymes.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -!- INDUCTION: Induced by heat shock during exponential growth and by
 CC heterologous amylases at the transition phase of the growth cycle.
 CC Negatively regulates its own expression.
 CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
 CC of *htrA*, especially during stress conditions.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to
 CC frameshifts in positions 87 and 246 that produce two separate
 CC ORFs.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF188296; AAF03153.1; .
 DR EMBL; Z93941; CAB07968.1; ALT_FRAME.
 DR EMBL; Z93941; CAB07969.1; ALT_FRAME.
 DR EMBL; Z93120; CAB15230.1; ALT_FRAME.
 DR Subtilist; BG4155; *yvtA*.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00889; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Heat shock; Transmembrane;
 KW Complete proteome.
 FT DOMAIN 1 71 CYTOSOLASMIC (POTENTIAL).
 FT TRANSMEM 72 92 POTENTIAL.
 FT DOMAIN 93 458 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 356 440 PDZ.
 FT ACT SITE 187 187 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 217 217 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 298 298 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 458 AA; 48717 MW; 77551045865A5CD CRC64;
 Query Match 19.5%; Score 350.5; DB 1; Length 458;
 Best Local Similarity 27.6%; Pred. No. 8.3e-13;
 Matches 111; Conservative 67; Mismatches 153; Indels 71; Gaps 16;
 QY 5 RRSRLWSLVSLAAV---GGLATAP-----AQAAPPALSDRFADFPALPLDPS 53
 DB 66 KKKKRAAWLSPILGIGTGGGLMGLIAPYPSDQNTQATETASANKVQSDNFTAPTNA 125
 QY 54 A---MVAQVAPQVNI-NTKLGVYNAV-----GAGTGIVI---DPNGVLTN 93
 DB 126 SNIAADVEDLEPTIVGISNIQTSQNTFTGGGSSSESSTGSGVIFPKDSKRAYITN 185

QY 94 NHVIAGATDINAFSVSGGTYGVVVGYDRTQVAVLQLRGAGGLPSAIGGG--VAVGE 151
 DB 186 NHVVEGANKLTV-TLYNGETAKLVGSDTITDLAVLEISGKNVKKVAFSGDSQRTGE 244
 QY 152 PTVAMGNSGGGTPRAVPRVVALGQTQASDLSLTGAETLNGLIQFPAALPGDGGGP 211
 DB 245 KVIAIGNPLGGQFSGVTQGIISGLNRTIDV-DITQGTVM--NVLTQDALNPGNSGP 301
 QY 212 VVNLGQWGMVMTAANDNFQSGG---QGFAIPICQAMAIAGQIRSGGSPVHIGPTA 268
 DB 302 LINASGVIGIN-----SLKVSSEGVESLGFALPSNDVPIVDQLLQNKV-----DRP 350
 QY 269 FLGLGVVDNN-----GNGARVQVRVGSAPASASLGISTGVDVITAVDGA 310
 DB 351 FLGVQIMDSQVPEYQENTLGLFGDQLGKGVYVKEVQANSFAEKAGIKSESDIVKLNKG 410
 QY 311 PINSATAMADAL-NGHHPGDVTSVMQTKSGGTRTGNVTLAE 351
 DB 411 DVSESSADIRQLLYKDLKVGDKTIIQ-VLRKGTKILNATLTK 451
 RESULT 2
 DEGS_ECOLI
 ID DEGS_ECOLI STANDARD; PRT; 355 AA.
 AC P31137;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 GN DEGS OR HOB precursor (EC 3.4.21.-).
 CN DEGS OR HOB OR HHRH OR B3235 OR Z4594 OR ECS4108.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Bass S., Gu Q., Goddard A.;
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=96165272; PubMed=8576051;
 RA Waller P.R., Sauer R.T.;
 RT "Characterization of degO and degS, Escherichia coli genes encoding
 RT homologs of the DegP protease.";
 RL J. Bacteriol. 178:1146-1153(1996).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:12453-1474(1997).
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RN SEQUENCE FROM N.A.


```

DR PIR: JC6051; JC6051.
DR MEROPS; S01.274; ...
DR SWISS-2DPAGE; P39059; COLI.
DR EcoGene; EGI2612; degQ.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 455
FT DOMAIN 258 349
FT ACT_SITE 109 109
FT ACT_SITE 139 139
FT ACT_SITE 214 214
SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;

Query Match
Best Local Similarity 30.4%; Pred. No. 1.9e-11;
Matches 119; Conservative 61; Mismatches 117; Indels 94; Gaps 19;

14 LLSVLA-AVGLGL-ATAPAAAPPALSDQRFADFPALPDSAMVAQVAPQVNT--- 68
7 LLSALALSGLTSLASFQAVASIPQGVADQ-APLESL-----APMLEKVLPAVSVRVEGT 61
69 -----KLGYNNAV-----GAGTGVIDPN-GVLTNNHVIAGATDINAFS 107
62 ASQGGKIPFEKFKFGDDLPQAPQPFGLSGVLIINASKGYLTNNHVINQAOKI-SIQ 120
108 VSGGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAIGGG--VAVGEPVAMNGSGGGGT 165
121 LNDGREFDAKLIGSDQSDALLQTONPSKLTQIAADSDKLVRGDFAVAVGNPFLGQT 180
166 PRAVGRVVALGQTVQASDSLTGAE-ETLNGLIQFDRAIQPDGSGGPPVNGLGQVGNWT 224
181 --ATSGIVSALGR-----SGLNLEGFNIQTDAISINRSGGALLNLELGINT 230
225 AASDNFQLSQGGQ-----GFAIPIGQMAIA-----GOIRSGGSPVTHIGFTAFGLGV 274
231 AI-----LAPGGSGVIGFAIPSNMARTLAQQLIDFGEIKRG-----LLGIKG 273
275 VDNNGN-----GARVQVVGSAAPASLGIISTGDTAVDVGAPINSATAMADLN 323
274 TEMSADIKAFLNDVORGAFVSEVLPGSGSAGVKGAGDITISLNGKPLNSFAELSRIA 333
324 GHHPGDVISVNMQTKSGGTRTG-----NVTL 349
334 TTEPGT-----KVLGLLNGKPLEVEVTL 358

RESULT 4
DEGP SALTY
ID DEGP SALTY STANDARD; PRT; 475 AA.
AC P26982;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR STM0209.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]_TaxID=602;
RP SEQUENCE FROM N.A.
RC STRAIN=CS;

```

Db 128 VTNNVVNDASVIVK-QLSDGRKFDKAVGKDRSDIALIQONPKNLTAIKLADSDALR 186
Qy 149 VGBPVVAMGNSGGGGTPRAVPRVVALGQTVQASDSLTGAE-ETLNGLIQFDAALQPGD 207
Db 187 VGDYVAIGNPFGGLGET--VTSGIVSALGR-----SGLNVENYENFIQTDAAINRGV 236
Qy 208 SGGPVVNLGQVVMNTA--ASDNFQSQGQGFAP-----IGQMAIAGQIRSGGGS 259
Db 237 SGGALVNLNGELIGINTAILAPDGGNI---GIGFAIPSNMVKNTSQMVEYGVQVXRG--- 290
Qy 260 PTVHIGPTAFLGLGVNDNNGN-----GARVORVWGSAPAASLIGIGTGVDVITA 306
Db 291 -----ELGIMGTLSNELAKAMKYDAQRGAFVSQVMNSNAKAGIKAGDVITS 339
Qy 307 VDGAPINSATAMADALNGHHPDGVISVNMWTKSGGTRTGNVTLAE 351
Db 340 LNKGPISFAALRAQVGTMPVSGKISLG-LLREGKAITVNLLEQQ 383

RESULT 5
HTRA_LACHE
ID HTRA_LACHE STANDARD; PRT; 413 AA.
AC Q9Z4H7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (EC 3.4.21.-).
GN HTRA.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=53/7;
RX MEDLINE=99047559; PubMed=9829922;
RA Smeds A., Varmanen P.K., Palva A.M.;
RT "Molecular characterization of a stress-inducible gene from
Lactobacillus helveticus";
RL J. Bacteriol. 180:6143-6153(1998).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ005672; CAA06668.1; -
DR MEROPS; S01.273; -
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_SIC.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
DR HydroLase; Serine protease; Transmembrane.
KW TRANSMEM 20 40
FT DOMAIN 141 302
FT CATALYTIC.
FT PDZ.
FT ACT SITE 305 401
FT ACT SITE 150 150
FT ACT SITE 180 180
FT ACT SITE 258 258
FT ACT SITE 258 258
SQ SEQUENCE 413 AA; 42647 MW; B16B677991C88707 CRC64;

Query Match 17.8%; Score 321; DB 1; Length 413;
Best Local Similarity 31.3%; Pred. No. 3.2e-11;
Matches 89; Conservative 50; Mismatches 121; Indels 24; Gaps 8;
Qy 67 NTKLGNNAVAGAGTGIVI---DPNGVLTNNHVIAGATDINAFVSGSQTYGVGVYDR 123
Db 119 SSKNGKLETYSSEGGVVMKSKNGYIVTNNHVISGSDAVQL-LANGKTVNAKVVGKDS 177
Qy 124 TODVAVLQIRAGGIPSAIAGG--VAVGEPVAMGNSGGQGTTPRAVGRVWALGQTVQ 181
Db 178 TTDLAVLSDIKAVYVTTQAFGDSKHLAAGQTVIAGVSLGSEYASTVTOGIIAPARTIS 237
Qy 182 ASDSLTGAETLNGLIQFDAALQPGDSGGPVVNGLVGVVGMNT--AASDNFQSQGQGF 239
Db 238 TS---SGNQOT---VIQDDAALNFGNSGALVNSAGVIGNSMKLAOSSDGTSGVGNAP 291
Qy 240 AIPICQAMAIAGQIRSGG--GSPTVHIGPTAFLG-----LGVDNNGNGARVORVVG 289
Db 292 AIPSNVTVVNLVKKGKITRPPQLGVRVIALQGIPEGRSRLKIKSLKNGIYIAFVSR 351
Qy 290 SAPASLIGISTGDVITAVDGAIPNSATAMADALNGHHPDGVISV 333
Db 352 NGSANAGIKSGDVITKVDGKKVEDVASLHSLILYSHKVGDTVNV 395

RESULT 6
DEGP_ECOLI
ID DEGP_ECOLI STANDARD; PRT; 474 AA.
AC P09376; P15724; Created)
DT 01-MAR-1989 (Rel. 10, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR ECS0165.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htra gene of Escherichia
coli: a sigma 32-independent mechanism of heat-inducible
transcription";
RL Nucleic Acids Res. 16:10053-10067(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
the 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

Db 123 KGVVTHNNVVDNATVTKV-QLSDGRKFDAMVKGKPRSDIALIOIQNPKNLTAIKWADS 191
Qy 145 GGVAVGEFVVMGNSGGQGGTPRAVGRVVALGQTVQASDSLTAIE-ETLNGLIQFDDAI 203
Db 182 DALRVGDVTVAGNPFGLGET--VTSIGVSLGR-----SGLNAENYENFIQTDAAI 231
Qy 204 QPGDSGGFVNVGLQVVGWMTA--ASDNFOLSGQGGFAIP-----IGAMALAGQIRS 255
Db 232 NRGNSGALVNLNGELGINTAILAPDGGNI---GIGFAIPNNVKNLTSCMVEYGGVGR 288
Qy 256 GGGSPVTHIGFTAPLGLGVVDNNGN-----GARVORVVGSAFAASLGISTGD 302
Db 289 G-----ELGIMGTLSLAKAMKVDQAORGAFVQVLPNSSAAKAGIKAGD 334
Qy 303 VITAVDGAIPNSATAM 318
Db 335 VITSLNGKPISSFAAL 350

RESULT 7
HTRA_LACIA
ID HTRA_LACIA STANDARD; PRT; 408 AA.
AC Q9LA06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (EC 3.4.21.-) (HtraL).
GN HTRA OR LU2136.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403";
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: DEGRADES ABNORMAL EXPORTED PROTEINS. NEEDED FOR THE PRO-
CC PEPTIDE PROCESSING OF A NATURAL PRO-PROTEIN AND FOR MATURATION OF
CC A NATIVE PROTEIN. RESPONSIBLE FOR THE HOUSEKEEPING OF EXPORTED
CC PROTEINS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF155705; AAF61294.1; -;
CC EMBL; AE006442; AAK06234.1; -;
CC PIR; H86891; H86891.
CC MEROPS; S01.273; -;
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase S1.
CC InterPro; IPR001940; Peptidase_S1C.

DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PRO0089; trypsin; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Transmembrane; Complete proteome.
FT TRANSMEM 6 26
FT DOMAIN 88 284
FT ACT SITE 127 127
FT ACT SITE 157 157
FT ACT SITE 239 239
SQ SEQUENCE 408 AA; 41648 MW; 581B90E55A7DF851 CRC64;
Query Match 17.2%; Score 310; DB 1; Length 408;
Best Local Similarity 31.3%; Pred. No. 1.3e-10;
Matches 93; Conservative 51; Mismatches 119; Indels 34; Gaps 11;
Qy 79 GTGIVDPNG---VVLTHNVIAGATDINAFVSGQGVVGVVGYDTQDVAVLQLRG- 134
Db 108 GSGVIYKSGDAYVVTNVTNVIAGNSLDVLLSG-GQVKASVVGVDVTDLAVLKISSE 166
Qy 135 -AGGLPSAAGGVAVGPPVAVGNSGGGGTPRAVGRVVALGQTVQASDSLTAIEEFL 193
Db 167 HVKDVATFADSSKLTIGEPFAIAGVPLGSPFANTATEGILSATSRQVLTQE-NGQTNI 225
Qy 194 NGLIQFDAAIQDGGGPPVNVGLGVVGMNTA---ASDNFOLSGQGGFAIPFGQAMAIA 250
Db 226 NA-IQTDRAINFNSGGLINIEGVIGITQSKITTTEDGTSVEGLGFAIPSNVDVNI 284
Qy 251 GQIRGGGSPVTHIGFTAPLGLGVVD-----NNGN-----GARVORVVGSAFA 294
Db 285 NKLEADG-----KISRPA-LGIRVMDLSQLNSDSSQLKLPSSVTVGVVGVVQGLPAA 338
Qy 295 SIGISTGVDVITAVDGAIPNSATAMADALNGHHPGVVISVNWQTKSGGTRTGNVTLAE 351
Db 339 SAGLKAGDVITKVGDTAVTSSDLSALYSHNINTVKTYV-RDGKSNATADVKLKSK 394

RESULT 8
DEGP_BRUME STANDARD; PRT; 513 AA.
AC O8YG32;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR EMEI1330.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Andersen I., Bhattacharya A., Lykidis A., Renik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- SUBCELLULAR LOCATION: Periplasmic (potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF155705; AAF61294.1; -;
CC EMBL; AE006442; AAK06234.1; -;
CC PIR; H86891; H86891.
CC MEROPS; S01.273; -;
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase S1.
CC InterPro; IPR001940; Peptidase_S1C.

QY 71 -----GNNVAGAGTGIVIDPENGVLNNHVIAGATDINAF 106
DB 104 DFGMEPRGSDNRNRKANKRPRGFEREVAQSGFVYVYNNHW---SDGDAY 160
QY 107 SV--GSGQTYGVVDVYDRTQDAVLQRLRGAGLPS-----AAIG--GGVAVGEFVWANG 157
DB 161 TVVLDGTELDKALIGADPTDLAVLKINA-----PKRKEVYVAFGDDNKVYGVWVAVG 216
QY 158 NSGGQGTTPRAGRVVAGVQVQASDLSLTGABETLNGLIQDAAIQPDGSGPVVNGLG 217
DB 217 NPFGLGCT--VTSGIVSARGRDIGAG-----PYDDFQIDAAYVKNKSGGPAFDLSG 266
QY 218 QVGVNTASDAPFOLSGGOG--FATPIGAMAIAGQIERSGG-----GSTVTHIGP----- 266
DB 267 EVIGINTAI---PSPGGVGVIAFAIPSTAKQVVDQLIKKSVRGWIGVQIOPVTKDI 323
QY 267 TAFGLGVVDNNGNGARVORVVGSAASLHSTGDTITAVDAPINSATAMADALNGHH 326
DB 324 AASLGLA-----BEKGAIVASPDQDGAAXAGIKAGDVITAVNGETVQDPRDLARKVANIA 379
QY 327 PGDVISVNWTKSGGRTGNTVLAEGP 353
DB 380 PGKKAALTVMKKNKABEI-NVTIAAMP 405

RESULT 10
HTRA_BACSU STANDARD; PRT; 449 AA.
AC 034358;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease do-like htra (EC 3.4.21.-).
GN HTRA OR BSU12900.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Devine K.M.;
RL "Sequence of the Bacillus subtilis genome between xlyA and ykoR";
RT Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=38044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosoro S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara K., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
Sato T., Scanlan E., Schleich S., Schroter R., Scoffone F.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Taccori E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
Toato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis";
RL Nature 390:249-256(1997).
RN [3]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20158875; PubMed=10692364;
RA Noone D., Howell A., Devine K.M.;
RT "Expression of ykxA, encoding a Bacillus subtilis homologue of HtrA,
is heat shock inducible and negatively autoregulated";
RL J. Bacteriol. 182:1592-1599(2000).
RN [4]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20576168; PubMed=11133960;
RA Noone D., Howell A., Colley R., Devine K.M.;
RT "YkxA and ykxA, HtrA-like serine proteases in Bacillus subtilis,
engage in negative autoregulation and reciprocal cross-regulation of
ykxA and ykxA gene expression";
RL J. Bacteriol. 183:654-663(2001).
RN [5]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=21439741; PubMed=1155295;
RA Hytylaenen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P.,
Vitikainen M., Sarvas M., Pragai Z., Bron S., van Dijk J.M.,
Kontinen V.P.;
RT "A novel two-component regulatory system in Bacillus subtilis for the
survival of severe secretion stress";
RL Mol. Microbiol. 41:1159-1172(2001).
CC -!- FUNCTION: May be involved in processing, maturation, or secretion
of extracellular enzymes.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- INDUCTION: Transcription is cspS dependent. Induced by heat shock
during exponential growth and by heterologous amylases at the
transition phase of the growth cycle. Negatively regulates its own
expression during exponential growth and during heat shock. Expression
of ykxA, especially during stress conditions.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ002571; CAA05570.1; -;
DR PIR; A69643; A69643.
DR MEROPS; S01.273; -;
DR Subtilist; BG12608; htra.
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
KW Hydrolyase; Protease; Serine protease; Heat shock; Transmembrane;
Complete proteome.
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 67 POTENTIAL.
FT DOMAIN 68 449 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 103 108 POLY-SER.

FT DOMAIN 146 152 POLY-SER.
 FT DOMAIN 348 437 PDZ.
 FT ACT SITE 179 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 209 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 290 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 449 AA; 47713 MW; E12B07A9018E8414 CRC64;

Query Match 16.4%; Score 295.5; DB 1; Length 449;
 Best Local Similarity 29.3%; Pred. No. 8.9e-10;
 Matches 93; Conservative 53; Mismatches 126; Indels 45; Gaps 11;

QY 53 SAMVAQVQVNNINIKLGYNA-----VGAGTGIVI-DPNG--VLTNN 94
 Db 119 SDVEDLSPAIVGINTLQAQSSSLFGSSSSSDETSFGSGGVIFKENGKAVIITNN 178
 QY 95 HVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQRLGAGLPSAAIG--GGVAVGEP 152
 Db 179 HVVEGASSLKV-SLYDGTETAKLVGSDSLTDLAVLQISDDHVTKVANFGDSSDLRTGET 237
 QY 153 VVAMGNSGGGQTPRAVGRVVALQTVQASDLSLTGAEETLNGLIQFDAAIQPGDSGGPV 212
 Db 238 VIAIGDPLGKLSRTVTCIGVSGVDRTVSM-----TSAGETINVIQTDAAINPGNSGGL 294
 QY 213 VNLGLGVGMNT--AASDNFQLSGGGQFAIPIGQAMAIAGIQRSGGSPVHIGPTAFL 270
 Db 295 LNTDGKIVGNSMKISEDDVE---GIGFAIPSNVDVKPIAEELLSKGIERPYIG-VSML 349
 QY 271 GIGVVDNN-----GNGARVORVGSAPASLGLSTGDTVAVDGPINSATA 317
 Db 350 DLEQVPQNYQEGTLGLFQSLNKGVIYREVASGSPAERKAGLKAEDIIIGLKGKIDTGE 409
 QY 318 MADAL-NGHHPGDVTSV 333
 Db 410 LRNLYLKAKIGDTVEV 426

RESULT 11
 HTOA HAEIN STANDARD; PRT; 466 AA.
 AC P45129;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable periplasmic serine protease do/hhoA-like precursor
 DE (EC 3.4.21.-).
 GN H11259.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geohagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
 RL Science 269:496-512(1995).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA (PROTEASE DO) AND HHOA.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32805; AAC22906.1; --
 CC PIR; A64113; A64113.
 CC MEROPS; S01.274; --
 CC TIGR; H11259;
 CC InterPro; IPR009003; Cys_Ser_trypsin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001940; Peptidase_S1C.
 CC Pfam; PF00595; PDZ; 2.
 CC Pfam; PF00089; trypsin; 1.
 CC PRINTS; PR00834; PROTEASES2C.
 CC SMART; SM00228; PDZ; 2.
 CC PROSITE; PS0106; PDZ; 2.
 CC Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
 CC Complete: proteome.
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 466 PROBABLE PERIPLASMIC SERINE PROTEASE
 FT DO/HHOA-LIKE.
 FT DOMAIN 270 361 PDZ 1.
 FT DOMAIN 367 438 PDZ 2.
 FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 466 AA; 49434 MW; ED050A00047B5851 CRC64;

Query Match 16.4%; Score 295; DB 1; Length 466;
 Best Local Similarity 28.1%; Pred. No. 9.9e-10;
 Matches 108; Conservative 54; Mismatches 124; Indels 99; Gaps 14;

QY 17 VLAAGVLGLATAP-----AQAPPALSQDRFADFPALPLDPSAMVAQVQVNNINIK 69
 Db 7 VLNSIALGLSVLSTSFVAHVAQATLPSFVSEONSLAP-----MLEKVPQVAVTSLVE 58
 QY 70 -----LGYNNAVAGTGIVDPN-GVLTN 93
 Db 59 GKAKVDSRSPFLDDIPEEFKFFGDRFAEFQFGRGESKRNFRGLSGVVIINASKGYVLTN 118
 QY 94 NHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQRLGAGLPSA--AIGGVAVGE 151
 Db 119 NHVIAGDKITV-QLQDGRFKAUVKDESDIALVQLKPSNLTETKFDADSKLRVGD 177
 QY 152 PVVAMGNSGGGQTPRAVGRVVALQTVQASDLSLTGAE--TLNGLIQFDAAIQPGDSGG 210
 Db 178 FTVAIGNPFLGQT--VTSGIVSALGRS-----TGSDSGTYENYIQTDAAVNRGNSGG 228
 QY 211 PVVNGLGQVGNMNTAASDNFQLSGGQ--FAIPIGQAMAI-----GQIRSGGSPV 262
 Db 229 ALVNLNGELIGINTAI--ISPSGNGAGIAFAPSNQASNLVQQLIEFGQVRG----- 279
 QY 263 HIGTAFGLGLGVDDNNGN-----GARVORVGSAPASLGLSTGDTVAVDGP 311
 Db 280 -----LLIGKIGELNADLAKAFNVSAQGFVSEVLKPSAAEKAGLKAGDIITAMNGQK 333
 QY 312 INSATAMADALNGHHPGDVTSV 335
 Db 334 ISSFAEIRAKIATTGAGKEISLT 357

RESULT 12
 DEGP_RHIME STANDARD; PRT; 504 AA.
 AC Q52894;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).

[illegible]

FT DOMAIN 286 357 PDZ 1.
 FT DOMAIN 419 466 PDZ 2.
 FT ACT SITE 143 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 173 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 247 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 503 AA; 54114 MW; 6CD9F473282AF9E CRC64;

Query Match
 Best Local Similarity 16.1%; Score 290; DB 1; Length 503;
 Matches 113; Conservative 42; Mismatches 135; Indels 118; Gaps 16;

QY 22 GLGLATAPAAQ---APPALQSRFPADFPALPLDPSAMVAQVAVQVNTKLG----- 71
 DB 29 GSSLWTKAHANSVFSSLMQQQGFAD-----IVSQKPAVVSQVZKSKKEWF 78
 QY 72 -----YN-----NAVGAGTGIVDPNGV 90
 DB 79 FSDPFSTPCFDLPQHPKXFFQDFYARDKPSNKSLSORSHRLPIAPGSGFFISSDGI 138
 QY 91 LNNHVIAGATDINAFSVSGQTYGVYVDRDQVAVLQURGAGGLPSAAIG--GGVA 148
 DB 139 VTNHVISDGTSY-AVLLDDGTGLNKLIGDPRDLAVLKVKNEKRFYSYVDFGDSKLR 197
 QY 149 VGEPPVAMGSGGGGTFRVGRVAVLQVQASDSLTGAEEITNGLIQFDAALQFGDS 208
 DB 198 VGSWVVALGNPFLGGTVA--GIVSARGR-----DIGTG---VYDFIQIDAAVNRGNS 247
 QY 209 GPPVNGVLGVVGMNTAASDNFQLSQGGQ--FAIPIGQAMAIAGAIRSGGSGPTVHGP 266
 DB 248 CGPTFDLNGKVVGVNTAI--FSPSGNVGIAFAIPATANEVVQQL----- 291
 QY 267 TAFGLGLGVNNGNGARVQV-----VG-----SAPASLGISITGDVITA 306
 DB 292 ---IEKGLVQRGWLGVQVQTPKREISDSIGLKEAKGALITDPLKPAKAGIKAGDVIIS 348
 QY 307 VDGAPINSATAMADALNGHPGDVISVN-WQTKSGGTRTGNTVLAEGP 353
 DB 349 VNGEKINDVRDLAKRIANWSFGETVTLGW--KSGKSENIKVKLDSNP 394

RESULT 14

ID DEGI ARATH STANDARD; PRT; 437 AA.
 AC O22609; Q9LKB5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 GN DEGI OR DEGP OR AT3G27925 OR K16N12.18.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RX MEDLINE=98175982; PubMed=9507020;
 RA Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;
 RT "Identification and characterization of DegP, a serine protease
 associated with the luminal side of the thylakoid membrane.";
 RN J. Biol. Chem. 273:7094-7098(1998).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA MEDLINE=20363099; PubMed=10907853;
 RX Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
 RL and BAC clones.";
 RN DNA Res. 7:217-221(2000).
 RN [3]
 RP SEQUENCE OF 104-118.
 RC STRAIN=cv. Columbia;

RA Kieselbach T., Bystedt M., Schroeder W.P.;
 RL Submitted (JUL-2000) to Swiss-Prot.
 CC -!- FUNCTION: Serine protease that is required at high temperature.
 CC May be involved in the degradation of damaged proteins. In vivo,
 CC can degrade beta-casein.
 CC -!- ENZYME REGULATION: Inhibited by phenylmethylsulfonyl fluoride and
 CC O-phenanthroline.
 CC -!- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID
 CC MEMBRANE.
 CC -!- INDUCTION: By heat shock.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF028842; AAC39436.1; -;
 CC EMBL; AF000371; BAB02539.1; -;
 CC EMBL; AF001302; BAB02539.1; JOINED.
 CC MEROPS; S01.279; -;
 CC InterPro; IPR009003; Cys_Ser_trypsin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001940; Peptidase_S1C.
 CC Pfam; PF00595; PDZ; 1.
 CC Pfam; PF00889; trypsin; 1.
 CC PRINTS; PRO0834; PROTEASES2C.
 CC SMART; SM00228; PDZ; 1.
 CC PROSITE; PS0106; PDZ; 1.
 KW Hydrolyase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
 FT TRANSIT 1 ?
 FT TRANSIT 2 ?
 FT CHAIN 104 437
 FT DOMAIN 152 321
 FT DOMAIN 324 421
 FT ACT SITE 171 171
 FT ACT SITE 201 201
 FT ACT SITE 280 280
 FT CONFLICT 12 23
 FT CONFLICT 36 36
 FT CONFLICT 54 54
 FT CONFLICT 60 60
 FT CONFLICT 64 64
 FT CONFLICT 68 69
 FT CONFLICT 355 355
 FT CONFLICT 381 381
 FT CONFLICT 416 416
 SQ SEQUENCE 437 AA; 46213 NW; 1497BIAB3F5FF2A4 CRC64;
 Query Match 16.1%; Score 289.5; DB 1; Length 437;
 Best Local Similarity 28.4%; Pred No. 1.9e-09;
 Matches 111; Conservative 56; Mismatches 143; Indels 81; Gaps 16;
 QY 7 RSLRWSLLSLVLA-----VGLGLATAPAAQAAPP-----ALSQDRFADPFA----- 47
 DB 42 RSKRYFRILSKPLNDNNGDDGDTLLTTPESAVKPFLLCTSVALSFSLFAAPVESA 101
 QY 48 -----LPLDPSAMV---AQVAPQVNTKLGNNAV-----GAGTGVID 85
 DB 102 SAFVVSSTPKKLTDELATVRLFQENTSVVITNLAVRQDAFTLDVLEVPQSGSGFW 161
 QY 86 PNGVLTNNHVIAGATDINAFSVSGQTYGVYVDRDQVAVLQURGAGG-----LFSA 141
 DB 162 KQGHVITNHYHVRGASDLRV-TLADQTTFDKAVGVGFDQKDAVLRIDAPKPKLRIP-V 219
 QY 142 AIGGVAVGEPVAMGNSGGGGTFRVAVGRVAVLQVQASDSLTGAEEITNGLIQFDA 201

Db 220 GVSADLLVCQKVFATGNPFLDHT--LITGVISGLRREI--SSAATG--RPIQDVQIOTDA 273
QY 202 AIQPGDSGGVVGVLGVVGMNTAASDNFQLSQGGQFAIPGQAWAIAQIRSGGSGPT 261
Db 274 AINPNSGGLLDSSTLIGINTAIYSPGAS--SGVGFSPVDVTGIVDQL----- 324
QY 262 VHLGPTAFGLGV-----VDNNG--NGARVQVVGSAFAAALGIST-----GD 302
Db 325 VFGKVTIRILGKFAPODSVQLGVGVLVLDAPSPGAGKAGLQSTKRDGYGRLLIIGD 384
QY 303 VITAVDGPINSATAMADALNGHHPGDVIVS 333
Db 385 IITSVNGTKVNSGDIYRLDCKQGVDEVTV 415

RESULT 15

DEGP CHLTR STANDARD; PRT; 497 AA.
ID AC PI8584; O84830;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa
DE immunogenic protein) (SK59).
GN DEGP OR HTRA OR CT823.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar L2;
RX MEDLINE=9000809; PubMed=2379836;
RA Kahane S., Weinstein Y., Sarov I.;
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RT "Cloning, characterization and sequence of a novel 59-kDa protein of
RT Chlamydia trachomatis."
RL Gene 90:61-67(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Atavind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC -!- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
CC THE AUTHORS TRANSLATED THEIR PUTATIVE 59 kDa IMMUNOGENIC PROTEIN
CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL
CC PROTEIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AE001355; AAC68420.1; -.
EMBL; M31119; AAA23116.1; -.
PIR; H71465; H71465.
PHCI-2DPAGE; P18584; -.
InterPro; IPR009003; Cys_Ser_trypsin.
InterPro; IPR001478; PDZ
InterPro; IPR001254; PDZ
InterPro; IPR001254; Peptidase S1.
InterPro; IPR008256; Peptidase_S1B_v8.
InterPro; IPR001940; Peptidase_S1C.
Pfam; PF00595; PDZ; 2.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00834; PROTEASES2C.

PRINTS; PR00839; V8PROTEASE.
SMART; SM00228; PDZ; 2.
PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Repeat; Signal; Antigen;
KW Complete proteome.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 497 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 128 289 CATALYTIC.
FT DOMAIN 290 381 PDZ 1.
FT DOMAIN 394 485 PDZ 2.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 497 AA; 53244 MW; 86A5E31BB84A38BA CRC64;

Query Match 16.0%; Score 288; DB 1; Length 497;
Best Local Similarity 32.3%; Pred. No. 2.6e-09;
Matches 95; Conservative 46; Mismatches 109; Indels 44; Gaps 12;
QY 79 GTGIVDPNGVLTNNHVIAGTDINAFSGSGGTGVVGVYDRTQDVAVLQRGAGGL 138
Db 127 GTGIVSEGGVVTNNHVVEDAGKIHV--TLHDGQKYAKIVGLDPKTDLAVIKIQ--AEKL 184
QY 139 PSAAIGGG--VAVGEPVVMGNSGGGTTPRAVPRVVALGQT---VQASDSDLTGAET 192
Db 185 PFLTGNSDQLQIGDWAIAIGNPFLQAT--VTGVVISAKGRNQLHIVDFED----- 234
QY 193 LNGLIQFDAAIQDGSQGVVVGVLGVVGMNTAASDNFQLSQG--GQGFAPICQAMAIA 250
Db 235 ---FIQTDALNPGNSGGPLNINQVIGVNTAIVSG---SGYIGIGFAIPSLMAKRV 288
QY 251 QGISGGSPVTHIGPTAFGLGVVDN-----NNGARVQVVGSAFAAALGISTG 301
Db 289 DQLISDGQVTRGFLGVT---LQPIDSELATCYKLEKYGVGALVTDVWVGSPAELKGRQE 344
QY 302 DVITAVDGPINSATAMADALNGHHPGD--VISVNWQTKSGGTRTGNVTLAEGP 353
Db 345 DVIVAYNGKEVESLSALRNAISLMMPGTRVVKI---VREGKTIETPTVTQIP 395

Search completed: June 30, 2004, 16:49:41
Job time : 11.7798 secs

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 1, 1863. It is a very important document, as it contains the President's message to the Congress, and it is the first of its kind. The President's message is a very important document, as it contains the President's message to the Congress, and it is the first of its kind. The President's message is a very important document, as it contains the President's message to the Congress, and it is the first of its kind.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 25.8569 Seconds
(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796C-4

Perfect score: 1802

Sequence: 1 MSNRRRLSRWELLSVLAA.....QTKSGGTRGNVTLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_minc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1792	99.4	355	007175	007175 mycobacteri
2	1792	99.4	355	16 Q7U2S9	Q7U2S9 mycobacteri
3	1306	72.5	361	2 Q50320	Q50320 mycobacteri
4	1273.5	70.6	354	16 Q9CCY9	Q9CCY9 mycobacteri
5	469.5	26.1	445	16 Q8VKA4	Q8VKA4 mycobacteri
6	469.5	26.1	464	16 Q7U0X2	Q7U0X2 mycobacteri
7	469.5	26.0	464	16 Q53896	Q53896 mycobacteri
8	460	25.5	382	16 Q9CD67	Q9CD67 mycobacteri
9	460	25.5	452	2 Q9Z5G6	Q9Z5G6 mycobacteri
10	407	22.6	542	16 Q9FBK9	Q9FBK9 streptomyc
11	396	22.0	519	16 Q93J30	Q93J30 streptomyc
12	391	21.7	375	16 Q8DG87	Q8DG87 synchococc
13	382	21.2	394	16 P72780	P72780 synchocyst
14	382	21.2	472	16 Q82FM9	Q82FM9 streptomyc
15	382	21.2	619	16 Q82IL8	Q82IL8 streptomyc
16	379.5	21.1	473	16 Q8PR17	Q8PR17 corynebacte

17	373.5	20.7	500	2 Q9KJN6	Q9KJN6 myxococcus
18	372	20.6	407	16 Q8YTF9	Q8YTF9 anabaena sp
19	366	20.3	525	16 Q89QJ8	Q89QJ8 bradyrhizob
20	364.5	20.2	675	16 Q8G6T3	Q8G6T3 bifidobacte
21	363	20.1	395	16 Q7V5C8	Q7V5C8 prochloroco
22	362	20.1	525	16 Q8PMV4	Q8PMV4 xanthomonas
23	358	19.9	413	16 Q8NS10	Q8NS10 corynebacte
24	355.5	19.7	452	16 P73354	P73354 synchocyst
25	354.5	19.7	490	16 Q8XPT5	Q8XPT5 ralstonia s
26	354	19.6	371	16 Q31388	Q31388 bradyrhizob
27	354	19.6	433	16 Q97GD5	Q97GD5 clostridium
28	353	19.6	326	16 Q8DMV9	Q8DMV9 synchococc
29	349	19.4	447	16 Q8R756	Q8R756 thermoaer
30	347.5	19.3	391	16 Q893T5	Q893T5 clostridium
31	347.5	19.3	432	16 Q7U8K9	Q7U8K9 synchococc
32	347	19.3	374	16 Q7U495	Q7U495 synchococc
33	347	19.3	465	16 Q92Q66	Q92Q66 rhizobium m
34	347	19.3	525	16 Q8PB56	Q8PB56 xanthomonas
35	346	19.2	389	16 Q8DL28	Q8DL28 synchococc
36	345	19.1	339	16 Q89RP2	Q89RP2 bradyrhizob
37	345	19.1	362	16 Q9SZR5	Q9SZR5 streptomyc
38	344	19.1	511	2 Q8RTK2	Q8RTK2 xanthomonas
39	343	19.0	457	16 Q8ZB58	Q8ZB58 yersinia pe
40	343	19.0	463	16 Q8DIR0	Q8DIR0 yersinia pe
41	342.5	19.0	528	16 Q89321	Q89321 bradyrhizob
42	341.5	19.0	514	16 Q9PBA3	Q9PBA3 xylella fas
43	340.5	18.9	505	16 Q8Y0I6	Q8Y0I6 ralstonia s
44	338.5	18.8	629	16 Q7U0X4	Q7U0X4 rhodospirell
45	335.5	18.6	429	16 Q8YYZ0	Q8YYZ0 anabaena sp

ALIGNMENTS

RESULT 1

007175 PRELIMINARY; PRT; 355 AA.

AC 007175: (TREMELrel. 04, Created)

DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)

DT 01-JUL-1997 (TREMELrel. 25, Last annotation update)

DE Hypothetical protein (Serine protease, putative).

GN PEPA OR RV0125 OR MTCI418B.07 OR MT0133.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RA "Complete genome sequence."

RT Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J.F., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains."

RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

KW Hydrolase; Protease; Serine protease; Signal.
FT SIGNAL 1 38
SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD5F3C411 CRC64;
Query Match 72.5%; Score 1306; DB 2; Length 361;
Best Local Similarity 70.4%; Pred. No. 1.6e-61;
Matches 255; Conservative 41; Mismatches 58; Indels 8; Gaps 3;
QY 1 MNS-RRSRRLRSWLLSLVLAAYGLG-ATAPAAAPPALSDREDFPPLPLDPS 53
DB 1 MNSRRSRRLRSWLLSLVLAAYGLG-ATAPAAAPPALSDREDFPPLPLDPS 53
QY 54 AMVAQVAPQVNNITKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFVSGSGQT 113
DB 61 AMVQVGPQVNNITKFGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFVSGSGQT 120
QY 114 YGVVVYGVYDRTQVAVLQIRGAGLPSAALGGVAVGEPVWAMGNSGGGGTTPRAVPGRV 173
DB 121 YAVDVVYGVYDRTQVAVLQIRGAGLPSAALGGVAVGEPVWAMGNSGGGGTTPRAVPGRV 180
QY 174 VALGQTVQASDLSLTGAETLNGLIQDAATQPGDSGGPVVNGLGQVVMNTAASDNFOLS 233
DB 181 VALNQSVAISDLSLTGAETLNGLIQDAATQPGDSGGPVVNGLGQVVMNTAASDNFOLS 240
QY 234 QGQGFPAIPQGMALAGTQIRSGGGGPTVHIGTAFGLGVVDNNGNGARVQVWCSAPA 293
DB 241 -GGQGFPAIPGRAMAVANQIRSGAGSNTVHIGTAFGLGVVDNNGNGARVQVWCSAPA 299
QY 294 ASLGISGTVITAVDGPAPINSATAMADALNGHHPGDVIVSNWOTKSGGTRTGNVTLAEGP 353
DB 300 AARGIAPGDVITGVTPINGAISMTVELVPHHPGDTIAVFRSVDGGERTANITLAEGP 359
QY 354 PA 355
DB 360 PA 361
RESULT 4
Q9CCY9
ID Q9CCY9 PRELIMINARY; PRT; 354 AA.
AC Q9CCY9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable secreted serine protease.
GN ML2659.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=TN;
RY MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 405:1007-1011(2001).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583926; CAC32191.1; -;
DR PIR; A87242; A87242.
DR Leproma; ML2659; -;
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.

DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR008256; Peptidase S1B V8.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00083; trypsin; 1.
DR PRINTS; PRO0834; PROTEASES2C.
DR PRINTS; PRO0839; V8PROTEASE.
DR SMART; SMO228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9E4A CRC64;
Query Match 70.6%; Score 1272.5; DB 16; Length 354;
Best Local Similarity 70.1%; Pred. No. 9e-60;
Matches 251; Conservative 42; Mismatches 58; Indels 7; Gaps 3;
QY 1 MNSRRSRRLRSWLLSLVLAAYGLGATAPAAQ---APPALSDREDFPPLPLDPSAMVA 57
DB 1 MNSRRSRRLRSWLLSLVLAAYGLGATAPAAQ---APPALSDREDFPPLPLDPSAMVA-- 58
QY 58 QVAPQVNNITKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFVSGSGQTGYVD 117
DB 59 -VAPQVNNITKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFVSGSGQTGYVD 117
QY 118 VGVYDRTQVAVLQIRGAGLPSAALGGVAVGEPVWAMGNSGGGGTTPRAVPGRVVALG 177
DB 118 VGVYDRTQVAVLQIRGAGLPSAALGGVAVGEPVWAMGNSGGGGTTPRAVPGRVVALN 177
QY 178 QTVQASDLSLTGAETLNGLIQDAATQPGDSGGPVVNGLGQVVMNTAASDNFOLSQGG 237
DB 178 QTVQASEPLTGAETLNGLIQDAATQPGDSGGPVVNGLGQVVMNTAASDNFOLSQGG 236
QY 238 GFAIPQGMALAGTQIRSGGGGPTVHIGTAFGLGVVDNNGNGARVQVWCSAPAASLG 297
DB 237 GFAIPQGMALAGTQIRSGGGGPTVHIGTAFGLGVVDNNGNGARVQVWCSAPAASLG 296
QY 298 ISTGDISGTVITAVDGPAPINSATAMADALNGHHPGDVIVSNWOTKSGGTRTGNVTLAEGPPA 355
DB 297 ISVGLIISVDGPPISEATAMTNVLVPHHPGDTIAVFRSVDGGDLTANVTLAEGPPA 354
RESULT 5
Q8VKA4
ID Q8VKA4 PRELIMINARY; PRT; 446 AA.
AC Q8VKA4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Heat shock protein HtrA, putative.
GN MT1011.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gail J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006585; AAK45259.1; -;
DR TIGR; MT1011; -;
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.

QY 19 AAVGLGLATAPAAQAAPPALSDRFADFPALPLDPSA---MVAQVAPQVWNTKLYGNNNA 75
 Db 120 AASLGFENRAPGPGSGGVAASAAPSIPAANWPGSVEQVAAKVPSVVMLETDLGRQSE 179
 QY 76 VQAGTGVDPNGVVLNNHVIAGT-----DINAFSVSGQCTGVGVDRPDQV 127
 Db 180 --EGSGIILSAEGLITNNHVIAAAKPPLGSPPPKTTVTFSDDGTAPFTVVGADPTSDI 237
 QY 128 AVLQRLGAGGLPSAAIGGG--VAVGEPVVMGNSGGGCTPRAVPGRVVWALGQTVQASDS 185
 Db 238 AVRVQGVSGLTPISLGSSDLRVGQPVLAIGSPLEGST--VTGIVSALNRPVSTGE 295
 QY 186 LFGAETLNGLIQFPAATOPGDSGPGVNGLGQVVGNTA-----ASNFLSQS--GQGF 239
 Db 296 -AGNQNTVLDAIQTDAAINPGNSGALVNMNAQLVGNVSAIATLGDADSAQSGSILGF 354
 QY 240 AIPIGQMAIAQIRSGGSPVHIGTAPFLGLGVV--DNNNGARVORVVGSAAPASLGI 298
 Db 355 ALPVOQAKRIADELLSTGKA-----SHASLGQVQVNDKDLGAKIVEVVGAGNAAGV 408
 QY 299 STGDVITAVDGAPINSATAMADALNGHHPGDVSVNWTGSGGTRTGNVTLAE 351
 Db 409 PKGVVVTKVDDRPINSADALVAARSKAPGATVATFDPSGGSRVTVQVTLGK 461

RESULT 8
 Q9CD67 PRELIMINARY; PRT; 382 AA.
 AC Q9CD67
 DT 01-JUN-2001 (TREMELrel. 17, Created)
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Possible secreted serine protease.
 GN ML0176.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garner T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies K.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
 RA Holtzoy S., Hornsby T., Jagels S., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Butler S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AL583917; CAC29684.1; -;
 DR PIR; H86930; H86930.
 DR Leproma; ML0176; -;
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000903; Cys Ser trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 382 AA; 37084 MW; 3DB8DDDBAE32A80D CRC64;

Query Match 25.5%; Score 460; DB 16; Length 382;

Best Local Similarity 36.4%; Pred. No. 7.3e-17;
 Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;
 QY 22 GLGLATAQAAPALSDRFADFPALPLDPSAMVAQA---PQVYNINKLYGNNNAV 77
 Db 50 GAGPVTGAAASVPAANM-----PSGVEQVAVKVPVSVVMLETDLGRQSE-- 94
 QY 78 AGTGVDPNGVVLNNHVIAGA-----TDINAFSVSGQCTGVGVDR 123
 Db 95 EGSVILSADGLITNNHVAARSKAPGPGGLSPKTTVTF--DGTASFTVVGADP 151
 QY 124 TQDVAVLQRLGAGGLPSAAIGGG--VAVGEPVVMGNSGGGCTPRAVPGRVVWALGQTVQ 181
 Db 152 TSDIAVVRVQISGLTPTITGSSADLRVGPVAVGSPGLAGT--VTGIVSALNRPVS 209
 QY 182 ASDSLTGAETLNGLIQFPAATOPGDSGPGVNGLGQVVGNTA-----SDNFLSQS 235
 Db 210 TTGE--SGNQNTVLDAIQTDAAINPGNSGALVNMGQLVGNVSAIATLGDADSGDAQSGSI 268
 QY 236 GQGFAPITGQAMATAGIRSGGSPVHIGTAPFLGLGVVNDNGN--CARVQRVVSAPAA 294
 Db 269 GLGFAIPVQAKRIADELLISTG--KATH-----ASLGQVQVATDKGTPGAKVNDVVGAGAA 322
 QY 295 SLGISTGDVITAVDGAPINSATAMADALNGHHPGDVSVNWTGSGGTRTGNVTLAE 351
 Db 323 NAAVEKGVVLTKVDDRLJSSADALVAARSKAPGDKVSLTVQDSGSGSRVTVQVTLGK 379

RESULT 9
 Q9Z5G6 PRELIMINARY; PRT; 452 AA.
 AC Q9Z5G6
 DT 01-MAY-1999 (TREMELrel. 10, Created)
 DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Putative serine protease.
 GN MLCB373.28.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RL Harris D., Taylor K.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93188700; PubMed=8446027;
 RA Eiglmeyer K., Honore N., Woods S.A., Caudron B., Cole S.T.;
 RT "Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae."
 RL Mol. Microbiol. 7:197-206(1993).
 DR EMBL; AL035500; CAB36690.1; -;
 DR PIR; T45448; T45448.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000903; Cys Ser trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 452 AA; 45130 MW; 6CA675EB0911F983 CRC64;

Query Match 25.5%; Score 460; DB 2; Length 452;
 Best Local Similarity 36.4%; Pred. No. 9e-17;
 Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGIATAPAAQAAPALPSQORFADFPALPLDPPAMVAQVA-----POVYNINKLGNNAVG 77
 DB 120 GAGFVTGFAASVFAANN-----PSGSEVQAVKVPVSMVLETDLGRQSE-- 164

QY 78 AGTGIVDPNGVLTNNHVIAGA-----TDINAFSVGSGQTVGVVGYDR 123
 DB 165 EGSVILSADGLITNNHVVAVAAKGGGPGGLSPKTVTFE-----DRTASFTVVGADP 221

QY 124 TDQVAVLQLRGAGLPSAIAIGG--VAVGEPVVMNGSGGQGTTPRAVGRVVALGQTVQ 181
 DB 222 TSDIAVVRVQSIISGLTPITMGSSADLRVGPVAVGSPGLAGT--VTSGIVSALNRFVS 279

QY 182 ASDSLTGABETINGLIQFADAAIOPGDSGGPVVNGLQGVVGMATA-----SDNPLSQG 235
 DB 280 TTGE-SGNQTVLDAQTDAINPNSGALVNMGGOLVGNVSATLGAQSGAQSGSI 338

QY 236 GQGFAPIGQAMAIAGIIRSGSGSPVTHIGPTAFGLGLGVVDNNGN-GARVORVVGSAAPAA 294
 DB 339 GLGFAIPVDQAKRIADELISGT--KATH-----ASLGVQVATDKGTGPKAKVMDVWAGGAAA 392

QY 295 SLGISTGDVITAVDGPAPINSATAMALNGHHPGVDIVSNWOTKSGGTRTGNVTLAE 351
 DB 393 NAAVPGKGVLTVDKDDRLISSADALVAARSKAFGKGVSLTYQDSGSRVTVQVTLGK 449

RESULT 10

Q9P8K9 PRELIMINARY; PRT; 542 AA.

AC Q9P8K9
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SC05149 OR SCP8.12.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D., Harris D.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RA "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieffer H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

Hopwood D.A.;
 "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 Nature 417:141-147(2002).
 ENBL:AL939122; CAC01350.1; --
 DR GO:0008233; F:peptidase activity; IEA.
 DR GO:0004295; F:trypsin activity; IEA.
 DR GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro:IPR009003; Cys_Ser_trypsin.
 DR InterPro:IPR001478; PDZ.
 DR InterPro:IPR001254; Peptidase_S1.
 DR InterPro:IPR001940; Peptidase_S1C.
 DR Pfam:PF00595; PDZ; 1.
 DR Pfam:PF00889; trypsin; 1.
 DR PRINTS:PRO0834; PROTEASES2C.
 DR SMART:SM00228; PDZ; 1.
 DR PROSITE:PS0106; PDZ; 1.
 DR Hydroxylase; Protease; Serine protease; Complete proteome.
 KW SEQUENCE 542 AA; 53761 MW; 620F7D889DF7212D CRC64;
 SQ

Query Match 22.6%; Score 407; DB 16; Length 542;
 Best Local Similarity 32.6%; Pred. No. 7e-14;
 Matches 116; Conservative 56; Mismatches 138; Indels 46; Gaps 12;

QY 22 GLGIATAPAAQAAPPALPSQORFADFPALPLDPSAMVAQVAPVNNINKLGNNAVGAGTG 81
 DB 206 GVGTVLP--QAGPEAARD-----PDSVAGTAARALPSVTLH--VSGSEAGTGTG 254

QY 82 IVIDPENGVLNNHVI--AGATDINAFSVGSGQTVGVVGYDRVQVAVLQLRGAGLPL 139
 DB 255 FVLDGRGHILTNHVVPEAGSGEITVTFFNSGDTAAEAEEVWGRDGYLVAVKVGVTGLT 314

QY 140 SAAIGG--GVAVGEPVWAMNGSGGQTPRAVGRVVALGQTVQV--SDSLTGABETLNG 196
 DB 315 PMPGLNSQNVRVGDFVVAIGAPFLAGT--VTSGIISAKERPITAGGEGGSDISYVDA 372

QY 197 IQFDAAIQDSDGSGPVVNGLQGVVGMATA-----ASDNFOLSGGQGFAPICQAM 247
 DB 373 LQTDAPINFGSGGLLDARGAIGNSAIRSADSGSTESDDGQAGSIGLGFALPINQK 432

QY 248 ALAGAIRSGGSGPTVHIGPTAFGLGV--VDNN--GNGARVQ-----RVGSAAPASL 296
 DB 433 RVABEL-----INTGKAHPVIGITLDNNTYGDGARVSAGKGGDGPVTTGGPQAKA 484

QY 297 GISTGDVITAVDGPAPINSATAMALNGHHPGVDIVSNWOTKSGGTRTGNVTLAE 352
 DB 485 GIKPGDVITAVDQQRVHSGEELIVKTRAHPDRLELTQDRDGKTRVSLVLSGSSG 540

RESULT 11

Q93J30 PRELIMINARY; PRT; 519 AA.

AC Q93J30
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SC03977 OR SCBAC35E3.14.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Collins M., Harris D.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843435;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinaishi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";
RL Mol. Microbiol. 21:77-96(1996).
RN [4].
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Bown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)";
RL Nature 417:141-147(2002).
DR EMBL; AL939118; CAC4701.1; -;
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00200; Tryp_Spc; 1.
DR KQ Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 519 AA; 50327 MW; 5CB8D5F0CC19E428 CRC64;

Query Match 22.0%; Score 396; DB 16; Length 519;
Best Local Similarity 32.2%; Pred. No. 2.5e-13;
Matches 119; Conservative 54; Mismatches 148; Indels 48; Gaps 13;

QY 14 LLSVLAAGVLG-----LATAPQAAPPALSDQRFADFPALPLDPSAMVAQA 60
DB 159 LIAALVAGGLGGLGYTLARNNDSSGTTVSASDTGGSVKRDACTVA-----GVAAKAL 212
QY 61 PQVNTNKLGYNNVAGAGTIVDPNGVLTNNHVIAGATDINAFSV--GSGQTYGVDV 118
DB 213 PSTVTIQAB-GSNBEGGGTGFVDFKEGHIIVTNHVVAAVADGKLGATFPNGKKYDAEV 271
QY 119 VGYDRTQDVAVLQRLGA--GGLPSAAIGGG--VAYGEFVAMGNSGGGGTTPRAVPGRWVA 175
DB 272 VGHAGQYDVAVIKLENAPSLKPLALGSDSKVAVGDSITAIAGPFLSNT--VTTGIISA 329
QY 176 LGQTVQASDLSLTGAETLNGLIQFDALIQPDGSGPVNGLGVVGNNTA-----ASDNFQ 231
DB 330 KNRFPVSSDGSADSKASTMSALQTDASINPNSGGPLLDAGQNVGINSAIQSTGNGGFG 389
QY 232 LSQGGQ---GFAIPIGQAMAIAGQIRSGGSGPTVHIGTAFPLGLGVVDNNGGARV--QR 286
DB 390 TGGAGSIGLGFALPVNQAKFVAQQLIKSGKVPYAKIGASVSL-----EETNGAKLTEQG 444
QY 287 VVGS-----APASLSGISTGVTITAVDAGPINSATAMADALNCHHPGDVISVNWQKSG 340
DB 445 VGGSDPVEKGGPADDAGLKPGLVITKLLDRVIDSGPTLIGETWTHKFGDVEVTYIE-RGG 503
QY 341 GTRIGNVTIL 349
DB 504 KQHTAEVTL 512

RESULT 12
Q8DG87 PRELIMINARY; PRT; 375 AA.
AC Q8DG87;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Serine protease.
GN TLR2436.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1";
RL DNA Res. 9:123-130(2002).
DR EMBL; AF053377; BAC09988.1; -;
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
DR KQ Protease; Complete proteome.
SQ SEQUENCE 375 AA; 39724 MW; OD2602EB534915F5 CRC64;

Query Match 21.7%; Score 391; DB 16; Length 375;
Best Local Similarity 33.6%; Pred. No. 3.1e-13;
Matches 108; Conservative 49; Mismatches 114; Indels 50; Gaps 10;

QY 35 PALSODRFPAD-FPALPLDPSAMVAQVAVQVNVNTKLGYNNVAVGAGTIVDPNGVVLTN 93
DB 67 PLLSDPFFRFQFFPGLALPQE-----DLRGGSGFIIDPSGIVMIN 108
QY 94 NEVIAGATINAFSVGSGQTYGVVDVGYDRTQDVAVLQRLG-AGGLPSAAIGGG--VAYG 150
DB 109 AHVVSQADTVNV-RLKDGSRVFEVGVDEVSDLAIVLKGVTEPLTAPLGDSEVKVG 167
QY 151 EPPVAMGNSGGGGTTPRAVPGRWVAG--QTVQASDLSLTGAETLNGLIQFDALIQPDGDS 208
DB 168 DWAIAGVNPGLDNT-----VTLGIISTLHRSAAQVGIPDKRLDFIQTDAAINPGNS 219
QY 209 GGPVNVNGLGVVGMNTAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSGPTVHIG--P 266
DB 220 GGPLNLEAGEVIGINTA-----IRADWNGIGFAIPINKAKALQARLIRGEKIYHAYIGQM 275
QY 267 TAFPLGLGVVDNNG-----NGARVQRVVGSGAPASLSGISTGDTITAVDGAIPINSA 315
DB 276 TTFTPAMAXENNANPNPSVILPEVNVGLVQLVLPNTPAKAGLRWGDVITAVDGEPI TSA 335
QY 316 TAMADALNCHHPGDVISVNWQ 336
DB 336 DQLQTTIVDSAAVGVQVNLNLTVQ 356

RESULT 13
P72780

QY 213 VNLGVVGMNTAA-----SDNFQSQGQ---GFAIPGQMAIAGQIRSGGSPVTHIG 265
D5 468 LDKARVVGINSAIRSDSSDQSGAGSICLGFAPVNOAKRVAEELINTGRATHPVIG 527
QY 266 PTAFLGLGVVDNNGGARV-----QVVVGSAPAASLIGISTGDTAVDVGAPINGSATA 317
D5 528 VILDM-----DYTGEGARVGTYSNDGSPVTRGGPDRAGIQAGDVITEVDGQRIHSGEE 582
QY 318 MADALNGHHPGDVIVSNVQTKSGGTRTGN---VTLAEG 352
D5 583 LIVKIRAHPRGDRALATVE-----RDGKEKFTVLVLG 614

Search completed: June 30, 2004, 16:52:44
Job time : 30.8569 secs

QY 275 VDNNGGARVQVRVVGSAASLIGISTGDTAVDVGAPINSATAMADALNGHHPGDVISVN 334
D5 394 TEGCAGS--DAITPNPRAKAGLKPGDVTIKLDWVIDSGFTLGEIWTHERPGATVKLT 451
QY 335 WQTKSGGTRTGNVTIAE 351
D5 452 Y-TRDGKARTTDTVLGE 467

RESULT 15

Q82IL8 PRELIMINARY; PRT; 619 AA.
AC Q82IL8;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Putative serine proteinase.
GN SAV3115.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinoe M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: Deducing the ability of producing secondary
RT metabolites";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608305; PubMed=1262562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinoe M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.,
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL: AP005033; BAC70826.1;
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0007242; P:intracellular signaling cascade; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR01478; PDZ.
DR InterPro: IPR001254; Peptidase S1.
DR InterPro: IPR001940; Peptidase_S1C.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SMO0228; PDZ; 1.
DR Complete proteome.
SQ SEQUENCE 619 AA; 64251 MW; D2D3C72A67D3B3A6 CRC64;

Query Match 21.2%; Score 382; DB 16; Length 619;
Best Local Similarity 33.7%; Pred. No. 1.7e-12;
Matches 114; Conservative 53; Mismatches 127; Indels 44; Gaps 12;

QY 38 SODRPAFPALPLDPSAMVAQVAVQVYNINIKLYNNNAVAGTGIVIDPVGVLNNHVI 97
D5 298 SEERAADSVA-----GIAARALPSVTLHVK--GSAEAGTGIGFVLDGRGHILTNHVV 349
QY 98 --AGATDINAFSVSGQTYGVVDVGYDRTQDVAVLQRLGAGLPSAAIGG--GVAVGEFV 153
D5 350 EPAGSSGEISVTFSGGETAKATVVRGDSGYDLAVVYGVGGGLKPMPLGNSDNVQVGDV 409
QY 154 VAMGNSGGGGTTPRAVGRVVALGQTVQA--SDSITGAEETLNGLIQFDAAIQGDSGSPV 212
D5 410 VAIGAPFDLANT--VTSIIISAKERPITAGGKGGSDSVYDALQTDAPINFGNSGGL 467

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very important document, as it contains the President's annual message to Congress. The letter is written in a formal, dignified style, and it is one of the most important documents in the history of the United States. It is a document that has been read and studied by many generations of Americans, and it is a document that has shaped the course of our nation's history.

2. The second part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very important document, as it contains the President's annual message to Congress. The letter is written in a formal, dignified style, and it is one of the most important documents in the history of the United States. It is a document that has been read and studied by many generations of Americans, and it is a document that has shaped the course of our nation's history.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 40.2057 Seconds
(without alignments)
2747.774 Million cell updates/sec

Title: US-09-597-796c-8

Perfect score: 1949

Sequence: 1 MVDFGALPEINSARMYAGP.....SGVLVPPRPYVMPHPAAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1949	100.0	391	2 AAW32381	Aaw32381 Mycobacte
2	1949	100.0	391	2 AAW32449	Aaw32449 Mycobacte
3	1949	100.0	391	2 AAW64335	Aaw64335 Mycobacte
4	1949	100.0	391	2 AAW81702	Aaw81702 M. tuberc
5	1949	100.0	391	2 AAY04778	Aay04778 Mycobacte
6	1949	100.0	391	2 AAY38989	Aay38989 M. tuberc
7	1949	100.0	391	2 AAY39132	Aay39132 M. tuberc
8	1949	100.0	391	4 AAU01888	AAU01888 M. tuberc
9	1949	100.0	391	5 AAE29707	AAe29707 Mycobacte
10	1949	100.0	391	5 AAE17571	AAe17571 Mycobacte
11	1949	100.0	596	2 AAY32070	AAy32070 Mycobacte
12	1949	100.0	596	5 AAE29710	AAe29710 Mycobacte
13	1949	100.0	596	5 AAE17574	AAe17574 Mycobacte
14	1949	100.0	599	5 AAU74599	AAU74599 Antigenic
15	1949	100.0	600	5 AAY32068	AAy32068 Mycobacte
16	1949	100.0	600	5 AAU74597	AAU74597 Antigenic
17	1949	100.0	723	7 ADA26354	Ada26354 Mycobacte
18	1949	100.0	729	4 AAO22142	AAO22142 Ra12-H9-3
19	1949	100.0	729	5 AAE29709	AAe29709 Mycobacte
20	1949	100.0	729	5 AAE17573	AAe17573 Mycobacte
21	1949	100.0	729	7 ADA26374	Ada26374 Mycobacte
22	1949	100.0	744	4 AAU01902	AAU01902 M. tuberc
23	1949	100.0	813	7 ADA26367	Ada26367 Mycobacte
24	1949	100.0	815	4 AAU01904	AAU01904 M. tuberc
25	1949	100.0	825	7 ADA26366	Ada26366 Mycobacte

26 1949 100.0 875 7 ADA26365
27 1949 100.0 930 5 AAE29731
28 1949 100.0 930 7 ADA26364
29 1949 100.0 1010 7 ADA26356
30 1949 100.0 1016 7 ADA26370
31 1949 100.0 1022 7 ADA26369
32 1949 100.0 1154 7 ADA26368
33 1945 99.8 788 4 AAU01903
34 1944 99.7 394 2 AAY04779
35 1944 99.7 729 5 AAE29708
36 1944 99.7 729 5 AAE17572
37 1939 99.5 729 7 ADA26373
38 1931 99.1 729 2 AAY32059
39 1902.5 97.6 726 5 AAY74588
40 1652.5 84.8 396 2 AAW64337
41 1652.5 84.8 396 2 AAW81704
42 1652.5 84.8 396 2 AAY38991
43 1652.5 84.8 396 2 AAY39134
44 1583 81.2 393 6 ABU36968
45 1583 81.2 393 6 ABU34531

ALIGNMENTS

RESULT 1
AAW32381
ID AAW32381 standard; protein; 391 AA.
XX
AC AAW32381;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TBH-9FL.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9709429-A2.
XX
PD 13-VAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-VAR-1996; 96US-00620280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
XX (CORI-) CORIXA CORP.
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedwick TH, Twardzik DE;
XX
DR WPI; 1997-192904/17.
DR N-PSDB; AAT91455.
XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
XX useful for diagnosis of M. tuberculosis infection.
PS Example 3; Page 150-152; 190pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,
CC TBH-9FL The immunogenic polypeptide can be used to diagnose
CC M.tuberculosis infection by forming complexes with specific antibodies in
CC the sample. Fragments of DNA encoding the immunogenic polypeptide can be
CC used as diagnostic primers or probes and agents that bind to the antigen,

Ada26365 Mycobacte
Aae29731 Mycobacte
Ada26364 Mycobacte
Ada26356 Mycobacte
Ada26370 M. bovis
Ada26369 Mycobacte
Ada26368 Mycobacte
AAU01903 M. tuberc
AAY04779 Mycobacte
AAE29708 Mycobacte
AAE17572 Mycobacte
ADA26373 Mycobacte
AAY32059 Mycobacte
AAU74588 Antigenic
AAW64337 Mycobacte
AAW81704 M. tuberc
AAY38991 M. tuberc
AAY39134 M. tuberc
ABU36968 Protein e
ABU34531 Protein e

PR 13-MAR-1997; 97US-00818111.
 XX (CORI-) CORIXA CORP.
 XX
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 XX Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-251292/22.
 DR N-PSDB; AAV44395.
 XX
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 XX
 XX Example 3; Page 133-135; 250pp; English.
 XX
 XX This polypeptide comprises Mycobacterium tuberculosis antigen TBH-9FL. It
 CC is encoded by genomic DNA (see AAV44395) isolated from a M. tuberculosis
 CC strain H37Rv genomic library using a probe from clone TBH-9 (see
 CC AAV44371). The invention relates to compositions and methods for
 CC diagnosing tuberculosis. It provides polypeptides (see AAV44391-W64379)
 CC comprising an antigenic portion of a soluble M. tuberculosis antigen, or
 CC an immunogenic portion of an M. tuberculosis antigen, as well as DNA
 CC sequences encoding such polypeptides, recombinant expression vectors and
 CC transformed or transfected host cells. Also claimed are methods and
 CC diagnostic kits for detecting M. tuberculosis infection in a patient
 CC using these polypeptides, antibodies or oligonucleotide probes and
 CC primers, for the diagnosis of tuberculosis. (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142; Indels 0; Gaps 0;
 Matches 391; Conservative 0; Mismatches 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSGWIG 60
 DB 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSGWIG 60
 QY 61 SSAGLMVAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
 DB 61 SSAGLMVAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
 DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180
 QY 181 LLEQAAVEEASDPTAAANQLMNNVPQALQQAQTPGTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAVEEASDPTAAANQLMNNVPQALQQAQTPGTPSSKLGGLWKTVPSPHRSPI 240
 QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
 DB 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
 QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
 DB 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
 QY 361 QMGARAGGSLGVLVPPRPYVMPHSPAAG 391
 DB 361 QMGARAGGSLGVLVPPRPYVMPHSPAAG 391

RESULT 4

AAW81702
 ID AAW81702 standard; protein; 391 AA.

XX AC AAW81702;

XX DT 27-JAN-1999 (first entry)

XX

DE M. tuberculosis immunogenic polypeptide TBH-9FL.
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 XX Mycobacterium tuberculosis.
 XX
 XX WO9816646-A2.
 PN 23-APR-1998.
 PD
 XX 07-OCT-1997; 97WO-US018293.
 PF
 XX 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00818112.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64503.
 XX
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 XX Example 3B; Page 128-129; 230pp; English.
 PS
 XX This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis
 XX
 XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142; Indels 0; Gaps 0;
 Matches 391; Conservative 0; Mismatches 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSGWIG 60
 DB 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSGWIG 60
 QY 61 SSAGLMVAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
 DB 61 SSAGLMVAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180
 DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180
 QY 181 LLEQAAVEEASDPTAAANQLMNNVPQALQQAQTPGTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAVEEASDPTAAANQLMNNVPQALQQAQTPGTPSSKLGGLWKTVPSPHRSPI 240
 QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
 DB 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
 QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
 DB 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
 QY 361 QMGARAGGSLGVLVPPRPYVMPHSPAAG 391
 DB 361 QMGARAGGSLGVLVPPRPYVMPHSPAAG 391

RESULT 5

AAV04778
 ID AAY04778 standard; protein; 391 AA.
 XX AC AAY04778;
 XX DT 06-JUL-1999 (first entry)
 XX MYcobacterium species protein sequence 5R.
 XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX MYcobacterium sp.
 XX WO9909186-A2.
 XX 25-FEB-1999.
 XX 14-AUG-1998; 98WO-FR001813.
 PR 14-AUG-1997; 97ER-00010404.
 XX 11-SEP-1997; 97ER-00011325.
 XX (INSP) INST PASTEUR.
 XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 XX WPI: 1999-181045/15.
 DR N-PSDB; AAX34030.
 XX MYcobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX Claim 32; Fig 5R; 309pp; French.
 XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX Sequence 391 AA;
 SQ
 Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 DB 1 MVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 QY 61 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
 DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
 QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 QY 241 MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSLGSSLGSSG 300
 DB 241 MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSLGSSLGSSG 300
 QY 301 LGGVAAANLGAASVGSLSVPOAAANOAVTPAARALPLTSLTSAERPGQWGLGLPV 360
 DB 301 LGGVAAANLGAASVGSLSVPOAAANOAVTPAARALPLTSLTSAERPGQWGLGLPV 360

QY 361 GQMGARAGGSLGVLRVPRPYPVMPHSPAAG 391
 DB 361 GQMGARAGGSLGVLRVPRPYPVMPHSPAAG 391
 RESULT 6
 AAY38989
 ID AAY38989 standard; protein; 391 AA.
 XX AC AAY38989;
 XX DT 05-NOV-1999 (first entry)
 XX M. tuberculosis recombinant antigen protein Tbh-9FL.
 XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX MYcobacterium tuberculosis.
 XX WO9942118-A2.
 XX 26-AUG-1999.
 XX 17-FEB-1999; 99WO-US003265.
 PR 18-FEB-1998; 98US-00024753.
 PR 05-MAY-1998; 98US-00072596.
 XX (CORI-) CORIXA CORP.
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX WPI: 1999-527416/44.
 DR N-PSDB; AA219093.
 XX New polypeptide comprising antigenic portions of M. tuberculosis.
 XX Example 3; Page 168-169; 323pp; English.
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX Sequence 391 AA;
 SQ
 Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 DB 1 MVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 QY 61 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAAAAPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLTPFEAPEMTSAGG 180
 DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLTPFEAPEMTSAGG 180
 QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 QY 241 MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSLGSSLGSSG 300

```
Db 241 MVSNNHMTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 7
AAU01888
ID AAY39132 standard; protein; 391 AA.
AC AAY39132;
XX 05-NOV-1999 (first entry)
DE M. tuberculosis antigen TbH-9FL amino acid sequence.
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX Mycobacterium tuberculosis.
XX WO9942076-A2.
XX 26-AUG-1999.
XX 17-FEB-1999; 99WO-US003268.
XX 18-FEB-1998; 98US-00025197.
PR 05-MAY-1998; 98US-00072967.
XX (CORI-) CORIXA CORP.
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX WPI; 1999-527409/44.
DR N-PSDB; AAZ19305.
XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
PT tests and protective or therapeutic vaccines or compositions.
XX Example 3; Page 123-124; 299pp; English.
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
CC polypeptides fragments, can be used in pharmaceutical compositions or
CC vaccines to generate a protective or therapeutic immune response to M.
CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
CC killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249
CC to AAZ19460 and AAY39083 to AAY39225 are used in the exemplification of
CC the present invention
XX Sequence 391 AA;
SQ Query Match 100.0%; Score 1949; DB 2; Length 391;
Best Local Similarity 100.0%; Pred No. 1.2e-142; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0;

QY 1 MVDFGALPPPEINSGARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 1 MVDFGALPPPEINSGARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVTAENRAELMI 120
```

```
Db 61 SSAGLMVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVTAENRAELMI 120
QY 121 LIATNLLIGONTPATAVNEABYGEKWAODAAAMFGVAAATATATATLLPFEEAPEMTSAGG 180
Db 121 LIATNLLIGONTPATAVNEABYGEKWAODAAAMFGVAAATATATATLLPFEEAPEMTSAGG 180
QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPISN 240
Db 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPISN 240
QY 241 MVSNNHMTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
Db 241 MVSNNHMTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 8
AAU01888
ID AAU01888 standard; protein; 391 AA.
XX AAU01888;
XX 29-AUG-2001 (first entry)
DE M. tuberculosis antigen TbH9 (Mtb39A).
XX TbH9; Mtb39A; antigen; vaccine; tuberculosis; AIDS;
KW acquired immunodeficiency disease.
XX Mycobacterium tuberculosis.
XX WO200124820-A1.
XX 12-APR-2001.
XX 10-OCT-2000; 2000WO-US028095.
XX 07-OCT-1999; 99US-0158338P.
PR 07-OCT-1999; 99US-0158425P.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
PI WPI; 2001-290576/30.
DR N-PSDB; AAS03779.
XX Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens.
XX Example 2; Page 151-152; 168pp; English.
XX The sequence represents Mycobacterium tuberculosis TBH9 (also known as
CC Mtb39A), an M. tuberculosis antigen. Compositions comprising at least 2
CC heterologous antigens, as a fusion protein, and vectors expressing the
CC fusion proteins are used as vaccines to prophylactically immunise mammals
CC (especially humans) against infection by Mycobacteria. The compositions
CC contain at least 2 heterologous antigens that increase the serological
CC sensitivity of individuals infected with tuberculosis, a disease
CC frequently affecting patients with acquired immunodeficiency disease,
CC AIDS
XX Sequence 391 AA;
SQ Query Match 100.0%; Score 1949; DB 4; Length 391;
```

Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 61 SSAGLMVAAAAPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLMVAAAAPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLPPFEAPMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLPPFEAPMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVQTAQNGVRAMSSLGSSLGSSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVQTAQNGVRAMSSLGSSLGSSG 300

QY 301 LGGGVAANLGRAASVGSLSVFOAWAANAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVFOAWAANAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 360

QY 361 GOMGARAGGSLSGVLRVPPRPVMPHSPAAG 391
DB 361 GOMGARAGGSLSGVLRVPPRPVMPHSPAAG 391

RESULT 9
AAE29707
ID AAE29707 standard; protein; 391 AA.
AC AAE29707;
XX
XX
DT 27-JAN-2003 (first entry)
XX
DE Mycobacterium sp. TbH9FL antigenic protein.
XX
XX
KW Vaccine; immunity; diagnostic agent; gene therapy; TbH9FL antigen.
XX
OS Mycobacterium sp.
XX
FN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Brannon M, Guderian J;
XX
XX WPI; 2002-759844/82.
DR N-PSDB; AAD47082.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.

Disclosure; Page 86-87; 155pp; English.

The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected

CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention are used in methods for eliciting immune response in mammals. They are useful as vaccines to elicit protective immunity against pathogenic microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion polypeptides are used for enhancing the expression of polynucleotides, as in vivo diagnostic agents and for raising antibodies in a non-human animal. The invention is used in gene therapy. The present sequence is CC Mycobacterium sp. TbH9FL antigenic protein
XX
SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 5; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 61 SSAGLMVAAAAPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLMVAAAAPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLPPFEAPMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLPPFEAPMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVQTAQNGVRAMSSLGSSLGSSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVQTAQNGVRAMSSLGSSLGSSG 300

QY 301 LGGGVAANLGRAASVGSLSVFOAWAANAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVFOAWAANAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 360

QY 361 GOMGARAGGSLSGVLRVPPRPVMPHSPAAG 391
DB 361 GOMGARAGGSLSGVLRVPPRPVMPHSPAAG 391

RESULT 10

AAE17571
ID AAE17571 standard; protein; 391 AA.

AAE17571;
AC
XX
XX
DT 22-APR-2002 (first entry)
XX

DE Mycobacterium species MTB39 (TbH9) protein #2.

KW Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB39; TbH9 protein.

OS Mycobacterium sp.

XX WO200198460-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US019959.

XX 20-JUN-2000; 2000US-00597796.

XX 01-FEB-2001; 2001US-0265737P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Alderson M;

XX WPI; 2002-147798/19.

DR N-PSDB; AAD28341.
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 PS Claim 83; Page 102-103; 136pp; English.
 XX
 CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB39 (TBH9) protein
 XX
 SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 5; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 60
 DB 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 60
 QY 61 SSAGLMVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
 DB 61 SSAGLMVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
 DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
 QY 181 LLEQAAVVEEASDTAAANQLMNNVPAALQQLAQTQGTTPSSKLGGLWKTVPSPHPSISN 240
 DB 181 LLEQAAVVEEASDTAAANQLMNNVPAALQQLAQTQGTTPSSKLGGLWKTVPSPHPSISN 240
 QY 241 MYSMANNHMTNSGVSWTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGSSG 300
 DB 241 MYSMANNHMTNSGVSWTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGSSG 300
 QY 301 LGGVVAANLGRASVGLSVPOQAWAANQAVTPAARALPLTSLTSAARPGQMLGGLPV 360
 DB 301 LGGVVAANLGRASVGLSVPOQAWAANQAVTPAARALPLTSLTSAARPGQMLGGLPV 360
 QY 361 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 391
 DB 361 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 391

RESULT 11
 AAY32070
 ID AAY32070 standard; protein; 596 AA.
 XX
 AC AAY32070;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen fusion protein Mtb59f.

XX Tuberculosis; antigen; fusion protein; Mtb59f; TBH9; Ra35; diagnosis;
 KW therapy; vaccine; immunogen.
 XX Mycobacterium tuberculosis.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..8
 FT Protein /note= "Met/His tag"
 FT Protein 9..140
 FT Protein /note= "Ra12"
 FT Protein 143..596
 FT Protein /note= "TBH9"
 XX WO9511748-A2.
 XX 14-OCT-1999.
 PD
 XX 07-APR-1999; 99WO-US007717.
 PF
 XX 07-APR-1999; 98US-00056556.
 PR
 XX 30-DEC-1999; 98US-00223040.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky YAW, Alderson M, Campos-Neto A;
 XX WPI; 1999-601610/51.
 PI N-PSDB; AAZ20205.
 DR
 XX New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.
 PT Claim 1; Fig 12A-B; 83pp; English.
 PS
 XX This sequence represents a recombinant Mycobacterium tuberculosis bi-
 CC antigen fusion protein, termed Mtb59f, composed of the antigens TBH9 and
 CC Ra35. The fusion protein is expressed in host cells using a vector
 CC carrying a polynucleotide (see AAZ20205) comprising the coding sequences
 CC for the 2 antigens. The invention provides fusion proteins (see AAY32059-
 CC 71) containing at least 2 M. tuberculosis antigens. The new fusion
 CC proteins and polynucleotides encoding them are useful as vaccines for
 CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
 CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),
 CC monitoring of disease progression, and treatment of tuberculosis. They
 CC are more effective immunogens than mixtures of the individual protein
 CC components
 CC
 XX Sequence 596 AA;
 SQ

Query Match 100.0%; Score 1949; DB 2; Length 596;
 Best Local Similarity 100.0%; Pred. No. 2.1e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 60
 DB 9 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 68
 QY 61 SSAGLMVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
 DB 69 SSAGLMVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
 QY 181 LLEQAAVVEEASDTAAANQLMNNVPAALQQLAQTQGTTPSSKLGGLWKTVPSPHPSISN 240
 DB 189 LLEQAAVVEEASDTAAANQLMNNVPAALQQLAQTQGTTPSSKLGGLWKTVPSPHPSISN 248
 QY 241 MYSMANNHMTNSGVSWTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGSSG 300
 DB 249 MYSMANNHMTNSGVSWTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGSSG 308

QY 301 LGGVAAANLGRAASVGSLSVPOAAANAOVTPAARALPLTSLTSAABRGQOMLGGLPV 360
 Db 309 LGGVAAANLGRAASVGSLSVPOAAANAOVTPAARALPLTSLTSAABRGQOMLGGLPV 368
 QY 361 GOMGARAGGGLSGVLRVPRPYVMPHSPAAG 391
 Db 369 GOMGARAGGGLSGVLRVPRPYVMPHSPAAG 399

RESULT 12

AAE29710
 ID AAE29710 standard; protein; 596 AA.

XX AC AAE29710;

XX DT 29-AUG-2003 (revised)

XX DT 27-JAN-2003 (first entry)

XX DE Mycobacterium sp. MTB59F fusion protein.

XX KW Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen; Ra35;
 KW MTB59F; fusion protein.

XX OS Mycobacterium sp.

XX OS Mycobacterium tuberculosis.

XX OS Chimeric.

XX PN WC200272792-A2.

XX PD 19-SEP-2002.

XX PF 13-MAR-2002; 2002WO-US008223.

XX PR 13-MAR-2001; 2001US-0275837P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Brannon M, Guderian J;

XX DR WPI; 2002-759944/82.

XX DR N-PSDB; AAD47086.

XX PT New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif.
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.

XX PS Disclosure; Page 98-99; 155pp; English.

XX CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC MTB59F fusion protein. This fusion protein comprises Ra35 protein from
 CC Mycobacterium tuberculosis and TBH9 protein from Mycobacterium sp.
 CC (Updated on 29-AUG-2003 to standardise OS field)

XX SQ Sequence 596 AA;

Query Match

Best Local Similarity 100.0%; Score 1949; DB 5; Length 596;

Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

Db 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
 QY 61 SSAGLIVAAASPYVAMSVTAQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 69 SSAGLIVAAASPYVAMSVTAQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMAFGAAATATATATATALLPPEAPEMTSAGG 180
 Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMAFGAAATATATATALLPPEAPEMTSAGG 188
 QY 181 LLEQAAAABEASDTAAANQLMNNVPAALQQLAQTGGTTPSSKLGGLWKTVPSPHSPISN 240
 Db 189 LLEQAAAABEASDTAAANQLMNNVPAALQQLAQTGGTTPSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSILGSSLGSSG 300
 Db 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSILGSSLGSSG 308
 QY 301 LGGVAAANLGRAASVGSLSVPOAAANAOVTPAARALPLTSLTSAABRGQOMLGGLPV 360
 Db 309 LGGVAAANLGRAASVGSLSVPOAAANAOVTPAARALPLTSLTSAABRGQOMLGGLPV 368
 QY 361 GOMGARAGGGLSGVLRVPRPYVMPHSPAAG 391
 Db 369 GOMGARAGGGLSGVLRVPRPYVMPHSPAAG 399

RESULT 13

ID AAE17574

XX AC AAE17574 standard; protein; 596 AA.

XX DT 22-APR-2002 (first entry)

XX DE Mycobacterium species MTB59F fusion protein.

XX KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB59F; TBH9-Ra35 protein.

XX OS Mycobacterium sp.

XX PN WO200198460-A2.

XX PD 27-DEC-2001.

XX PF 20-JUN-2001; 2001WO-US019959.

XX PR 20-JUN-2000; 2000US-00597796.

XX PR 01-FEB-2001; 2001US-0265737P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Reed S, Alderson M;

XX DR WPI; 2002-147798/19.

XX DR N-PSDB; AAD28344.

XX PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.

XX PS Claim 5; Page 114-115; 136pp; English.

XX CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding such and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,

RESULT 15

AAV32068
ID AAV32068 standard; protein; 600 AA.
XX
AC AAV32068;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mycobacterium tuberculosis antigen fusion protein Mtb61f.
XX
KW Tuberculosis; antigen; fusion protein; Mtb61f; Tbh9; DPV; MTL; diagnosis;
KW therapy; vaccine; immunogen.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9951748-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US007717.
XX
PR 07-APR-1998; 98US-00056556.
PR 30-DEC-1998; 98US-00223040.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Alderson M, Campos-Neto A;
XX
DR WPI; 1999-601610/51.
DR N-PSDB; AAZ20203.
XX
PT New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis.
XX
PS Claim 1; Fig 10A-B; 83pp; English.
XX
CC This sequence represents a recombinant Mycobacterium tuberculosis tri-
CC antigen fusion protein, termed Mtb61f, composed of the antigens Tbh9, DPV
CC and MTL. The fusion protein is expressed in host cells using a vector
CC carrying a polynucleotide (see AAZ20203) comprising the coding sequences
CC for the 3 antigens. The invention provides fusion proteins (see AAY32059-
CC 71) containing at least 2 M. tuberculosis antigens. The new fusion
CC proteins and polynucleotides encoding them are useful as vaccines for
CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),
CC monitoring of disease progression, and treatment of tuberculosis. They
CC are more effective immunogens than mixtures of the individual protein
CC components
XX
SQ Sequence 600 AA;

Query Match 100.0%; Score 1949; DB 2; Length 600;
Best Local Similarity 100.0%; Pred. No. 2.1e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 9 MVDFGALPPEINSARMYAGPGASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
QY 61 SSAGIMVAASPYVWWSVTAGQELTAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 120
DB 69 SSAGIMVAASPYVWWSVTAGQELTAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 128
QY 121 LIATNLLGQNTPAIVNAEYGEWMAQDAAMFCYAAATATATATLIPFEAEPEMTSAGG 180
DB 129 LIATNLLGQNTPAIVNAEYGEWMAQDAAMFCYAAATATATATLIPFEAEPEMTSAGG 188
QY 181 LLEQAAAABEASDTAAANQLMNNVPQALQLOAQTQCTTPSSKLGGLWKTVPSPHRSPI 240
DB 189 LLEQAAAABEASDTAAANQLMNNVPQALQLOAQTQCTTPSSKLGGLWKTVPSPHRSPI 248
QY 241 MVSNNHMTNSGVNTMTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
DB 249 MVSNNHMTNSGVNTMTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 308

QY 301 LGGVVAANLGRAASVCSLSVPCAWAAANQAVTPAARALPLTSLTSAARGPQOMLGGLPV 360
DB 309 LGGVVAANLGRAASVCSLSVPCAWAAANQAVTPAARALPLTSLTSAARGPQOMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPYPVMPHSPAAG 391
DB 369 GQMGARAGGGLSGVLRVPPRPYPVMPHSPAAG 399

Search completed: June 30, 2004, 16:48:47
Job time : 41.2057 secs

Sequence 16, Appl
Sequence 12, Appl
Sequence 142, App
Sequence 142, App
Sequence 131, App
Sequence 15, Appl
Sequence 57, Appl
Sequence 92, Appl
Sequence 114, App
Sequence 114, App
Sequence 109, App
Sequence 114, App
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl

US-09-287-849-16
US-09-287-849-12
US-09-073-009-142
US-09-073-010-142
US-09-073-135A-131
US-09-073-009-15
US-09-073-010-15
US-08-311-731A-57
US-08-311-731A-208
US-09-050-739-92
US-08-818-112-114
US-08-818-111-109
US-09-056-556-114
US-09-072-596-109
US-09-072-567-114
US-08-095-734-2
US-08-444-623-2
US-08-471-869-2

28 604 31.0 710 4
29 604 31.0 856 4
30 603 30.9 423 4
31 603 30.9 423 4
32 424.5 21.8 943 4
33 424 21.8 141 4
34 424 21.8 141 4
35 381.5 19.6 204 4
36 377.5 19.4 208 4
37 317 16.3 371 4
38 314 16.1 368 3
39 314 16.1 368 4
40 314 16.1 368 4
41 314 16.1 368 4
42 314 16.1 368 4
43 261 13.4 1271 1
44 261 13.4 1271 2
45 261 13.4 1271 3

ALIGNMENTS

RESULT 1
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-VAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-107

Query Match 100.0%; Score 1949; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 11.8942 Seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796C-8
Perfect score: 1949
Sequence: 1 MYDFGALPPEINSARMYAGP.....SGVLVPPRPVYMPHSPHAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	100.0	391	3	US-08-818-112-107
2	1949	100.0	391	4	US-08-818-111-102
3	1949	100.0	391	4	US-09-056-556-107
4	1949	100.0	391	4	US-09-072-596-102
5	1949	100.0	391	4	US-09-072-967-107
6	1949	100.0	596	4	US-09-287-849-26
7	1949	100.0	600	4	US-09-287-849-22
8	1944	99.7	729	4	US-09-223-040-2
9	1944	99.7	729	4	US-09-287-849-2
10	1852.5	84.8	396	3	US-08-818-112-111
11	1852.5	84.8	396	4	US-08-818-111-106
12	1852.5	84.8	396	4	US-09-056-556-111
13	1852.5	84.8	396	4	US-09-072-596-106
14	1852.5	84.8	396	4	US-09-072-967-111
15	1486.5	76.3	359	3	US-08-818-112-109
16	1486.5	76.3	359	4	US-08-818-111-104
17	1486.5	76.3	359	4	US-09-056-556-109
18	1486.5	76.3	359	4	US-09-072-596-104
19	1486.5	76.3	359	4	US-09-072-367-109
20	1187.5	60.9	358	4	US-09-287-849-8
21	1187	60.9	263	3	US-08-818-112-91
22	1187	60.9	263	4	US-08-818-111-92
23	1187	60.9	263	4	US-09-056-556-91
24	1187	60.9	263	4	US-09-072-596-92
25	1187	60.9	263	4	US-09-072-967-91
26	766.5	39.3	400	4	US-09-073-009-126
27	766.5	39.3	400	4	US-09-073-010-126

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
Db 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLWVAASPYVWAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
Db 61 SSAGLWVAASPYVWAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPFEEAPEMTSAGG 180
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPFEEAPEMTSAGG 180
QY 181 LLEQAAAEEASDTAAANQNMNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
Db 181 LLEQAAAEEASDTAAANQNMNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
Db 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPOAQAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPOAQAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 360
QY 361 GQMGARAGGSLGVLRVPPRYVMPHSPAAG 391
Db 361 GQMGARAGGSLGVLRVPPRYVMPHSPAAG 391

RESULT 2

US-08-818-111-102
; Sequence 102, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond S.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.41706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-102

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
Db 1 MVDGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLWVAASPYVWAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
Db 61 SSAGLWVAASPYVWAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPFEEAPEMTSAGG 180
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPFEEAPEMTSAGG 180
QY 181 LLEQAAAEEASDTAAANQNMNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
Db 181 LLEQAAAEEASDTAAANQNMNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
Db 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPOAQAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPOAQAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 360
QY 361 GQMGARAGGSLGVLRVPPRYVMPHSPAAG 391
Db 361 GQMGARAGGSLGVLRVPPRYVMPHSPAAG 391

RESULT 3

US-09-056-556-107
; Sequence 107, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
TREATME

US-09-056-556-107

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMAAASPYVAMSVTAGOAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMAAASPYVAMSVTAGOAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTALIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
DB 121 LIATNLGONTALIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
QY 181 LLEQAAAVEASDPTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLMKTVPSPHSPI 240
DB 181 LLEQAAAVEASDPTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLMKTVPSPHSPI 240
QY 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSILGSSG 300
DB 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSILGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGSLGVLPRPVPPRYMHPSPAG 391
DB 361 GQMGARAGGSLGVLPRPVPPRYMHPSPAG 391

RESULT 4

US-09-072-596-102
; Sequence 102, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
US-09-072-596-102

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMAAASPYVAMSVTAGOAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMAAASPYVAMSVTAGOAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTALIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
DB 121 LIATNLGONTALIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
QY 181 LLEQAAAVEASDPTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLMKTVPSPHSPI 240
DB 181 LLEQAAAVEASDPTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLMKTVPSPHSPI 240
QY 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSILGSSG 300
DB 241 MVSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSILGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGSLGVLPRPVPPRYMHPSPAG 391
DB 361 GQMGARAGGSLGVLPRPVPPRYMHPSPAG 391

RESULT 5

US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967

;; FILING DATE: 05-MAY-1998
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Maki, David J.
;; REGISTRATION NUMBER: 31,392
;; REFERENCE/DOCKET NUMBER: 210121.411C9
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 107:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 391 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-09-072-967-107

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
Db 1 MYDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLWVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPAIIVNEAEYGEWMAQDAAMFYAAATATATATLPPFEAPEMTSAGG 180
Db 121 LIATNLLGONTPAIIVNEAEYGEWMAQDAAMFYAAATATATATLPPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDSTAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
Db 181 LLEQAAVEEASDSTAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 241 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSLSGSSG 300
Db 241 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSLSGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAABERGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAABERGQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
Db 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391

RESULT 6

US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556

;; PRIOR FILING DATE: 1998-04-07
;; PRIOR APPLICATION NUMBER: US 09/223,040
;; PRIOR FILING DATE: 1998-12-30
;; NUMBER OF SEQ ID NOS: 46
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO: 26
;; LENGTH: 596
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 100.0%; Score 1949; DB 4; Length 596;
Best Local Similarity 100.0%; Pred. No. 5.2e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
Db 9 MYDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 68
QY 61 SSAGLWVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 69 SSAGLWVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
QY 121 LIATNLLGONTPAIIVNEAEYGEWMAQDAAMFYAAATATATATLPPFEAPEMTSAGG 180
Db 129 LIATNLLGONTPAIIVNEAEYGEWMAQDAAMFYAAATATATATLPPFEAPEMTSAGG 188
QY 181 LLEQAAVEEASDSTAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
Db 189 LLEQAAVEEASDSTAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
QY 241 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSLSGSSG 300
Db 249 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSLSGSSG 308
QY 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAABERGQMLGGLPV 360
Db 309 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAABERGQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 7

US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAODAAAMFGYAAATATATATLLPFEAPBMTSAGG 180
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAODAAAMFGYAAATATATATLLPFEAPBMTSAGG 321
QY 181 LLEQAAAEEASDTAAANQLMNNVPQALQQAQPTGTPSSKLGGLWKTVPSPHRSPIIN 240
DB 322 LLEQAAAEEASDTAAANQLMNNVPQALQQAQPTGTPSSKLGGLWKTVPSPHRSPIIN 381
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 300
DB 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARERGQMLGLPV 360
DB 442 LGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARERGQMLGLPV 501
QY 361 GOMGARAGGSLGVLRVPRPYVMPHSPAAG 391
DB 502 GOMGARAGGSLGVLRVPRPYVMPHSPAAG 532

RESULT 10

US-08-818-112-111
; Sequence 111, Application US/08818112
; Patent No. 6290965
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 13-MAR-1997
; APPLICATION NUMBER: US/08/818,112
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-818-112-111

Query Match 84.8%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTGSWIG 60

DB 1 VVDFGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTGSWIG 60
QY 61 SSAGLVAAASPYVWMSVTAGQBELTAAQVRAAAAYETAYGLTVPPRVIAENAEELMI 120
DB 61 SSAGLVAAASPYVWMSVTAGQBELTAAQVRAAAAYETAYGLTVPPRVIAENAEELMI 120
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 180
QY 181 LLEQAAAEEASDTAAANQLMNNVPQALQQAQPTGTPSSKLGGLWKTVPSPHRSPIIN 240
DB 181 LLEQAAAEEASDTAAANQLMNNVPQALQQAQPTGTPSSKLGGLWKTVPSPHRSPIIN 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 296
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 299
QY 297 GSSGLGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARERGQMLG 356
DB 300 GSSGLGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARERGQMLG 359
QY 357 GLPVGOMGARAG--GGLSGVLVRPVPYVMPHSPAAG 391
DB 360 GLPLGOLTNSGGFGGVSNAUMPPRAYVMPRVPAAG 396

RESULT 11

US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 13-MAR-1997
; APPLICATION NUMBER: US/08/818,111
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-818-111-106

Query Match 84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 VVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 61 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180

QY 181 LLEQAAVVEEASDTAAANQLMNNVPQALQQAQPTQGTSSKLGGLWKTVPSPHRSPLSN 240
DB 181 LLEQAAVVEEASDTAAANQLMNNVPQALQQAQPTQGTSSKLGGLWKTVPSPHRSPLSN 240

QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSS----LGSSL 296
DB 241 IVSMLNNHVSMTSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSS----LGSSL 299

QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLG 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLG 359

QY 357 GLPVGQMGARAG--GGLSGVLVRPPRYVMPHSPAAAG 391
DB 360 GLPLQLTNSGGGFGVSNALRMPFRAYVMPRVPAAG 396

RESULT 12
US-09-056-556-111
; Sequence 111, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-09-056-556-111

Query Match 84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 VVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 61 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180

QY 181 LLEQAAVVEEASDTAAANQLMNNVPQALQQAQPTQGTSSKLGGLWKTVPSPHRSPLSN 240
DB 181 LLEQAAVVEEASDTAAANQLMNNVPQALQQAQPTQGTSSKLGGLWKTVPSPHRSPLSN 240

QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSS----LGSSL 296
DB 241 IVSMLNNHVSMTSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSS----LGSSL 299

QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLG 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLG 359

QY 357 GLPVGQMGARAG--GGLSGVLVRPPRYVMPHSPAAAG 391
DB 360 GLPLQLTNSGGGFGVSNALRMPFRAYVMPRVPAAG 396

RESULT 13
US-09-072-596-106
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031

TREATY

INFORMATION FOR SEQ ID NO: 106:

SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-106

Query Match 84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 1 MYDFGALPPEINSARMYAGPGSASIVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 VVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAAASPYVAMSVTAGOAELETAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMSVTAGOAELETAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTFAIVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEEAPEMTSAGG 180
DB 121 LIATNLLGONTFAIVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVAMSS ---LGSSL 296
DB 241 IVSMNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVAMSS ---LGSSL 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAARPGOMLG 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAARPGOMLG 356
QY 357 GLPVGMGARAG--GGLSGVLVRPVPYVPHSPAAAG 391
DB 360 GLPLGQLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396

RESULT 14

US-09-072-967-111
Sequence 111, Application US/09072967

Patent No. 6592877
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967

FILING DATE: 05-MAY-1998

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-967-111

Query Match 84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 1 MYDFGALPPEINSARMYAGPGSASIVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 VVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAAASPYVAMSVTAGOAELETAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMSVTAGOAELETAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTFAIVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEEAPEMTSAGG 180
DB 121 LIATNLLGONTFAIVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVAMSS ---LGSSL 296
DB 241 IVSMNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVAMSS ---LGSSL 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAARPGOMLG 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAARPGOMLG 356
QY 357 GLPVGMGARAG--GGLSGVLVRPVPYVPHSPAAAG 391
DB 360 GLPLGQLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396

RESULT 15

US-08-818-112-109

Sequence 109, Application US/08818112

Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-818-112-109

Query Match      76.3%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.9e-116;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

Qy      1 MVDFGALPPPEINARMYAGSGASLVAAAKMWDVSDLFSAASAFQSVVMGLTVGSGWIG 60
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      1 VVDFGALPPPEINARMYAGSGASLVAAAKMWDVSDLFSAASAFQSVVMGLTVGSGWIG 60
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      61 SSAGLMVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      61 SSAGLMVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMT 120
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      121 LIATNLIGQNTPAIAVNEAYEGEMWQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 180
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      121 LTATNLIGQNTPAIEANQAAYSQWQDAEAMTYGAAATAATAATEALLPFEDAPLIITNPGG 180
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      181 LLEQAAVAVEASDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVSPHRSPISN 240
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      181 LLEQAAVAVEASDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVSPHRSPISN 240
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      241 MVSMAHHMNSMTNSGVSMNTTSLSMKGFAPAAAAQAVOTAAQNGVAMSS----LGSSL 296
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      241 VSSIANNHMSMMGTGVSMTNTLHSMKGLAP-AAAQAVETAENGVMAMSSSLGSLGSSL 299
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      297 GSSGLGCVAAANLGRAASVGSLSVPQNAWAANQAVTPAAPALPLTSLTSAERGPQMLG 356
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      300 GSSGLGAGVAAANLGRAASVGSLSVPPAWAANQAVTPAAPALPLTSLTSAATAPGHMLG 359
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Search completed: June 30, 2004, 16:55:19
Job time : 12.8942 secs

The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that proper record-keeping is essential for the integrity of the financial system and for the ability to detect and prevent fraud. The document also outlines the responsibilities of individuals involved in the process, including the need for transparency and accountability.

In addition, the document provides a detailed overview of the various methods used to collect and analyze data. It describes the different types of data sources, such as surveys, interviews, and focus groups, and explains how this information is used to identify trends and patterns. The document also discusses the importance of ensuring that the data is reliable and valid, and provides guidelines for how to conduct these activities effectively.

Finally, the document concludes by summarizing the key findings and recommendations. It highlights the need for continued research and monitoring to ensure that the system remains effective and efficient. The document also provides a list of resources and references for further information, and encourages stakeholders to work together to address any challenges that may arise.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:52:58 ; Search time 30.4893 Seconds
(without alignments)
3625.462 Million cell updates/sec

Title: US-09-597-796c-8

Perfect score: 1949
Sequence: 1 MYDFGALPPEINSAARMYAGP.....SGVLKVPVPRPYMHPSPAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	100.0	391	12	US-09-886-349A-14
2	1949	100.0	391	14	US-10-193-002-102
3	1949	100.0	391	14	US-10-084-843-107
4	1949	100.0	391	14	US-10-098-732A-14
5	1949	100.0	596	9	US-09-287-849-26
6	1949	100.0	596	12	US-09-886-349A-20
7	1949	100.0	596	14	US-10-359-460-26
8	1949	100.0	596	14	US-10-098-732A-20
9	1949	100.0	600	9	US-09-287-849-22
10	1949	100.0	600	14	US-10-359-460-22
11	1949	100.0	723	15	US-10-369-983-2
12	1949	100.0	723	15	US-09-886-349A-18
13	1949	100.0	729	14	US-10-098-732A-18
14	1949	100.0	729	15	US-10-369-983-21
15	1949	100.0	729	15	US-10-369-983-22

16	1949	100.0	813	15	US-10-369-983-15	Sequence 15, Appl
17	1949	100.0	825	15	US-10-369-983-14	Sequence 14, Appl
18	1949	100.0	875	15	US-10-369-983-13	Sequence 13, Appl
19	1949	100.0	930	14	US-10-098-732A-65	Sequence 65, Appl
20	1949	100.0	930	15	US-10-369-983-12	Sequence 12, Appl
21	1949	100.0	1010	15	US-10-369-983-4	Sequence 4, Appl
22	1949	100.0	1016	15	US-10-369-983-18	Sequence 18, Appl
23	1949	100.0	1022	15	US-10-369-983-17	Sequence 17, Appl
24	1949	100.0	1154	15	US-10-369-983-16	Sequence 16, Appl
25	1944	99.7	729	9	US-09-287-849-2	Sequence 2, Appl
26	1944	99.7	729	12	US-09-886-349A-16	Sequence 16, Appl
27	1944	99.7	729	14	US-10-359-460-2	Sequence 2, Appl
28	1944	99.7	729	14	US-10-098-732A-16	Sequence 16, Appl
29	1944	99.7	729	15	US-10-359-459-2	Sequence 2, Appl
30	1934	99.2	391	12	US-09-872-186-8	Sequence 8, Appl
31	1652.5	84.8	396	14	US-10-193-002-106	Sequence 106, App
32	1652.5	84.8	396	14	US-10-084-843-111	Sequence 111, App
33	1583	81.2	393	12	US-10-282-122A-62455	Sequence 62455, A
34	1583	81.2	393	12	US-10-282-122A-64892	Sequence 64892, A
35	1486.5	76.3	359	14	US-10-193-002-104	Sequence 104, App
36	1486.5	76.3	359	14	US-10-084-843-109	Sequence 109, App
37	1187.5	60.9	358	9	US-09-287-849-8	Sequence 8, Appl
38	1187.5	60.9	358	14	US-10-359-460-8	Sequence 8, Appl
39	1187	60.9	263	12	US-09-886-349A-12	Sequence 12, Appl
40	1187	60.9	263	14	US-10-193-002-92	Sequence 92, Appl
41	1187	60.9	263	14	US-10-084-843-91	Sequence 91, Appl
42	1187	60.9	263	14	US-10-098-732A-12	Sequence 12, Appl
43	766.5	39.3	400	9	US-09-073-009-126	Sequence 126, App
44	766.5	39.3	400	9	US-09-793-306-126	Sequence 126, App
45	742.5	38.1	405	12	US-10-282-122A-62027	Sequence 62027, A

ALIGNMENTS

RESULT 1

```
US-09-886-349A-14
; Sequence 14, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTE39 (TBH9FL)
US-09-886-349A-14
```

Query Match 100.0%; Score 1949; DB 12; Length 391;
Best Local Similarity 100.0%; Pred. No. 1e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MYDFGALPPEINSAARMYAGPGSASIVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB	1	MYDFGALPPEINSAARMYAGPGSASIVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY	61	SSAGLWVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPFVIAENRAELMI 120
DB	61	SSAGLWVAASPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPFVIAENRAELMI 120

REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 107:
US-10-084-843-107

Query Match
Best Local Similarity 100.0%; Score 1949; DB 14; Length 391;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

QY 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180

QY 181 LLEQAAVVEASDTAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAVVEASDTAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240

QY 241 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 300
DB 241 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360

QY 361 GQMGARAGGSLGVLVPPRPYVMPHSPAAG 391
DB 361 GQMGARAGGSLGVLVPPRPYVMPHSPAAG 391

RESULT 5
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
US-09-287-849-26

Query Match
Best Local Similarity 100.0%; Score 1949; DB 9; Length 596;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 241 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 300
Db 249 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 308
QY 301 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
Db 309 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 8

US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: Leishmania Antigen
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (designated MTB59P)
US-10-098-732A-20

Query Match 100.0%; Score 1949; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
QY 61 SSAGLVAAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
Db 69 SSAGLVAAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
QY 121 LIATNLLGQNTPAIAVNEAEYEMWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGQNTPAIAVNEAEYEMWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANQLMNNVPOALQOAOPTGTPSSKLGGLWKTVPSPHRSPI 240
Db 189 LLEQAAAVEASDTAAANQLMNNVPOALQOAOPTGTPSSKLGGLWKTVPSPHRSPI 248
QY 241 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 300
Db 249 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 308
QY 301 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
Db 309 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 10

US-10-359-460-22
; Sequence 22, Application US/10359460

RESULT 9
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-0090200S
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 100.0%; Score 1949; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
QY 61 SSAGLVAAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
Db 69 SSAGLVAAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
QY 121 LIATNLLGQNTPAIAVNEAEYEMWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGQNTPAIAVNEAEYEMWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANQLMNNVPOALQOAOPTGTPSSKLGGLWKTVPSPHRSPI 240
Db 189 LLEQAAAVEASDTAAANQLMNNVPOALQOAOPTGTPSSKLGGLWKTVPSPHRSPI 248
QY 241 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 300
Db 249 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 308
QY 301 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
Db 309 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

Publication No. US20030147911A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 22
LENGTH: 600
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-22

Query Match 100.0%; Score 1949; DB 14; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60
Db 9 MVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 68
QY 61 SSAGLWVAASPYVAMSVTAGQAEATAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 69 SSAGLWVAASPYVAMSVTAGQAEATAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAVEASDTPAAANQNMNVPAQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 240
Db 189 LLEQAAVEASDTPAAANQNMNVPAQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 248
QY 241 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLSGSSG 300
Db 249 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLSGSSG 308
QY 301 LGGGVAANLGRAASVGSLSVQPAWAAANQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 309 LGGGVAANLGRAASVGSLSVQPAWAAANQVTPAARALPLTSLTSAERGPQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 399

RESULT 11
US-10-369-983-2
Sequence 2, Application US/10369983
Publication No. US2003023593A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Reed, Steven
APPLICANT: Alderson, Mark
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009070US
CURRENT APPLICATION NUMBER: US/09/886,349A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: US 09/597,796
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: US 60/265,737
PRIOR FILING DATE: 2001-02-01
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.1

APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
CURRENT FILING DATE: 2003-02-18
PRIOR APPLICATION NUMBER: US 60/357,351
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 723
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: mutated
OTHER INFORMATION: MTE332-MTE39F fusion protein (MTE332MutSA)
US-10-369-983-2

Query Match 100.0%; Score 1949; DB 15; Length 723;
Best Local Similarity 100.0%; Pred. No. 2.3e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60
Db 333 MVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 392
QY 61 SSAGLWVAASPYVAMSVTAGQAEATAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 393 SSAGLWVAASPYVAMSVTAGQAEATAQVRVAAAAYETAYGLTVPPVIAENRAELMI 452
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 453 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 512
QY 181 LLEQAAVEASDTPAAANQNMNVPAQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 240
Db 513 LLEQAAVEASDTPAAANQNMNVPAQLAQPTQGTTPSSKLGGLWKTVPSPHRSPIIN 572
QY 241 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLSGSSG 300
Db 573 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLSGSSG 632
QY 301 LGGGVAANLGRAASVGSLSVQPAWAAANQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 633 LGGGVAANLGRAASVGSLSVQPAWAAANQVTPAARALPLTSLTSAERGPQMLGGLPV 692
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
Db 693 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 723

RESULT 12
US-09-886-349A-18
Sequence 18, Application US/09886349A
Publication No. US20040086523A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Reed, Steven
APPLICANT: Alderson, Mark
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009070US
CURRENT APPLICATION NUMBER: US/09/886,349A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: US 09/597,796
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: US 60/265,737
PRIOR FILING DATE: 2001-02-01
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.1

1 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Db 142 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 201
QY 61 SSAGLMVAASAPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 202 SSAGLMVAASAPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 180
Db 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAAVEASDTAAANQIMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHRSPI 240
Db 322 LLEQAAAVEASDTAAANQIMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHRSPI 381
QY 241 MVSMMNNHSMVNTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
Db 382 MVSMMNNHSMVNTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 441
QY 301 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 442 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532

RESULT 14
US-10-369-983-21
; Sequence 21, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F
US-10-369-983-21

Query Match 100.0%; Score 1949; DB 15; Length 729;
Best Local Similarity 100.0%; Pred. No. 2.4e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Db 142 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 201
QY 61 SSAGLMVAASAPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 202 SSAGLMVAASAPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 180
Db 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAAVEASDTAAANQIMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHRSPI 240
Db 322 LLEQAAAVEASDTAAANQIMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHRSPI 381
QY 241 MVSMMNNHSMVNTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
Db 382 MVSMMNNHSMVNTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 441
QY 301 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 442 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532

1 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Db 142 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 201
QY 61 SSAGLMVAASAPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 202 SSAGLMVAASAPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 180
Db 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAAVEASDTAAANQIMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHRSPI 240
Db 322 LLEQAAAVEASDTAAANQIMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHRSPI 381
QY 241 MVSMMNNHSMVNTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
Db 382 MVSMMNNHSMVNTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 441
QY 301 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 442 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532

RESULT 13
US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
US-10-098-732A-18

Query Match 100.0%; Score 1949; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 2.4e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Db 142 MVDGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 201
QY 61 SSAGLMVAASAPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 202 SSAGLMVAASAPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 180
Db 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAAVEASDTAAANQIMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHRSPI 240
Db 322 LLEQAAAVEASDTAAANQIMNNVPOALQQLAQTGTTTSSKLGGLWKTVPSPHRSPI 381
QY 241 MVSMMNNHSMVNTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
Db 382 MVSMMNNHSMVNTSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 441
QY 301 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 442 LGGVVAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532

Search completed: June 30, 2004, 17:14:44
Job time : 31.4893 secs

```
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
Db 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAARSGPQMLGGLPV 360
Db 442 LGGGVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAARSGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532
```

```
RESULT 15
US-10-369-983-22
; Sequence 22, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 22
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB72FmutSA (McB72f-mutSA)
US-10-369-983-22
```

```
Query Match 100.0%; Score 1949; DB 15; Length 729;
Best Local Similarity 100.0%; Pred. No. 2.4e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 142 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201
QY 61 SSAGLWAAASPYVAVWSVTAGQAEATAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 202 SSAGLWAAASPYVAVWSVTAGQAEATAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
QY 121 LIATNLGQNTPAIVNEAYGEMWADAAAFGYAAATATATATATLPPPEAPEMTSAGG 180
Db 262 LIATNLGQNTPAIVNEAYGEMWADAAAFGYAAATATATATLPPPEAPEMTSAGG 321
QY 181 LLEQAAVEASDPAANQALMNNVPQALQLOAQPTQCTTPSSKLGKWKTVSHRSPISN 240
Db 322 LLEQAAVEASDPAANQALMNNVPQALQLOAQPTQCTTPSSKLGKWKTVSHRSPISN 381
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
Db 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAARSGPQMLGGLPV 360
Db 442 LGGGVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAARSGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 10.2189 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796C-8

Perfect score: 1949
Sequence: 1 MVDGALPPEINARMYAGP.....SGVLVPRPRVMPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: Pirl:*

2: Pirl:*

3: Pirl:*

4: Pirl:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1949	100.0	391	2 B70608	probable PPE prote
2	1566.5	85.0	396	2 B70741	probable PPE prote
3	1583	81.2	393	2 C70568	probable PPE prote
4	779.5	40.0	393	2 G70329	probable PPE prote
5	751	38.5	409	2 A70332	probable PPE prote
6	737.5	37.8	403	2 H70331	probable PPE prote
7	736	37.8	423	2 B70331	probable PPE prote
8	731	37.5	421	2 H87056	PPE-family protein
9	705	36.2	408	2 G70325	probable PPE prote
10	702	36.0	391	2 B70825	probable PPE prote
11	702	36.0	413	2 F70560	probable PPE prote
12	688.5	35.3	463	2 C70331	probable PPE prote
13	682.5	35.0	468	2 B70332	probable PPE prote
14	675	34.6	380	2 A70646	probable PPE prote
15	668.5	34.3	394	2 G70881	probable PPE prote
16	667.5	34.2	385	2 H70303	probable PPE prote
17	635	32.6	350	2 H70329	probable PPE prote
18	633.5	32.5	365	2 E70329	probable PPE prote
19	615.5	31.6	402	2 C70882	probable PPE prote
20	603	30.9	423	2 C70582	probable PPE prote
21	597.5	30.7	391	2 D70322	probable PPE prote
22	589	30.2	406	2 E70875	probable PPE prote
23	588	30.2	394	2 A70504	probable PPE prote
24	575.5	29.5	391	2 A70663	probable PPE prote
25	495.5	25.4	3300	2 D70375	probable PPE prote
26	473	24.3	180	2 G70834	probable PPE prote
27	452.5	23.2	3716	2 E70969	probable PPE prote
28	448.5	23.0	580	2 G70570	probable PPE prote
29	444	22.8	346	2 H70874	probable PPE prote

ALIGNMENTS

RESULT 1

B70608

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70608

R:Coile, S.I.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70608

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-391 <COL>

A:Cross-references: GB:Z93777; GB:AL123456; NID:g3261726; PIDN:CAB07839.1; PID:e311073; I

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match	100.0%	Score 1949;	DB 2;	Length 391;
Best Local Similarity	100.0%;	Pred. No. 3.5e-108;		
Matches 391;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MVDGALPPEINARMYAGPGSALVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG	60	
DB	1	MVDGALPPEINARMYAGPGSALVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG	60	
QY	61	SSAGLWVAASPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPPEVIAENRAELMI	120	
DB	61	SSAGLWVAASPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPPEVIAENRAELMI	120	
QY	121	LIATNLLGQNTPAIAVNEAEYGEWQAQDAANFGVAAATATATATLLPFEAPENTSAGG	180	
DB	121	LIATNLLGQNTPAIAVNEAEYGEWQAQDAANFGVAAATATATATLLPFEAPENTSAGG	180	
QY	181	LLEQAAAEEASDSTAANQLMNNVPAQLQAQPTQGTTPSKLGLWKTVPSPHRSPLSN	240	
DB	181	LLEQAAAEEASDSTAANQLMNNVPAQLQAQPTQGTTPSKLGLWKTVPSPHRSPLSN	240	
QY	241	MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVQTAQNGVRAMSSLGSSGSG	300	
DB	241	MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVQTAQNGVRAMSSLGSSGSG	300	
QY	301	LGCGVAANLGRAASVGSLSVFCQAAANCAVTPAARALPLTSLTSAERGPQMLGGLPV	360	
DB	301	LGCGVAANLGRAASVGSLSVFCQAAANCAVTPAARALPLTSLTSAERGPQMLGGLPV	360	
QY	361	GQGARAGGSLGVLVRVPRPVMPHSPAAG	391	

122 IATNLLGONTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATATLLPFBEAPMTSAGGL 181
121 ISTNIFGONTSAIAAAEAQYGENWQAQDAAMFYAGSSASASA-VTFPFTPTPOLANPTAQ 179
182 LEQAAAVEREASDTAAA--NCLMNNVPCALQQLAQAOTGQTPSPSKLGLMK----- 229
180 GTQAAAVATAGTQAGTGLTETMTITGLPNAQLSTLSPLLQSS-NGPLSLWQLLPGTENFT 238
230 -----TVSPHSPISNNVSMANNHMTNNGSVSMNTLTLSSMLKGFAPAAAQAQVOTAA 282
239 SISALLTDLQPVASFFNYTEGLPVFSGMGNFIOAKTL-GLIGSAAPAFAVA-----AA 292
283 QNGVRAMSSILGSSGLGGVGAANLGRAASVGSLSVPOAWAA-ANQAVTPAARALPLT 341
293 GDAAKGLPLGLGMLG---GGPVAAGLGNRAASVGKLSVPPVNSGPLPGSVTFPGAAPLPVS 348
342 SLTSAAEPRGQMLGGLPVGQMGARAGGGLSGVLVPPPRPYNMPHSPAAG 391
349 TVSAAPAEAPGSLGGLLGLPL-----AGAGGAGAGP-RYGFRETVMARPPFAG 393
RESULT 5
A70932 probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: A70932
R/Colo, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A/Authors: Sgares, R.; Sultaton, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: A70932
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-409 <COL>
A/Cross-references: GB:AL022021; GB:AL123456; MID:g3250699; PIDN:CAA17729.1; PID:e125461
A/Experimental source: strain H37Rv
C/Genetics:
A/Gene: PPE
Query Match 39.5%; Score 751; DB 2; Length 409;
Best Local Similarity 44.7%; Pred. No. 2.2e-37;
Matches 177; Conservative 53; Mismatches 138; Indels 28; Gaps 10;
2 VDFGALPPEINSAWYAGPGSASLVAAAQWDSVASDLFSAASAFQSVWGLTVGSWIGS 61
1 MDFGALPPEINSGRMVAGPGSGPLLAALAAWDLAAELYSAAASYSGTIEGLTVAPWMPG 60
62 SAGILWAAAAPYVAWWSVTAGQALTAQVRAAAAYETAYGLTVPPPPVTAENRAELMIL 121
61 SSITMAAAAPYVAWISVTAGQAQAGQAKIANGVYETAFAATVPVPIEANKALLMSL 120
122 IATNLLGONTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATATLLPFBEAPMTSAGGL 181
121 VATNIFGONTPAIAATEAHVAEMWQAQDAAMFYAGSSATA-SQLAPFSPPQTNPFSAT 179
182 LEQAAAVEREASDTAAA-----NOLMNNVPCALQQLAQAOTGQTPSPSKLGLMKWTSVSPH 234
180 AAQGAHVQAQGAASDDITRAQLSQLISLLPSTLQSLA--TTATATASAG--WDTV--- 232
235 RSPISNNVSMANNHMTNNGSVS-----MTNLTSSMLKGFAPAAAQAQVOTAAQNGVRAM 289
233 LQSITITLANLTGPVSIIGLGAIPGGWMLTFQGLGAQNAQAPGVAALLGPKAAAGALSPL 292
290 SSL-GSSILGS-SGLGGVGAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSA- 346
293 APLRGYIGDITPLPGGATGATGIAIRIYVGSLSVPOGWAEEAPVWRAVASVLPGTGAAPAL 352
347 AERFGQMLGGLPVGQMGARAGGGL-----SGVLRV 377

Db 353 AAEAPGALFEGALSSLAGRALAGTAVRSGAARV 388

RESULT 6

H70931

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70931

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, K.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; PMID:98295987; PMID:9634230

A:Accession: H70931

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-403 <COL>

A:Cross-references: GB|AL020201; GB|AL123456; NID:G3250699; PIDN:CAA17728.1; PID:e125461

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

	Query Match	Score	DB 2;	Length	403;
	Best Local Similarity	42.6%;	Pred. No. 1.4e-36;		
	Matches 181;	Conservative	60;	Mismatches 123;	Indels 61; Gaps 12;

Qy	2	VDFGALPPEINSARMYAGPGSASLVAAQQWDSVASDLFSAASAQSVVVGLTVGSWGTS	61
Db	5	LDPATLPPEINSARMYSGAGSAPMLAAGAATHGLSAELRASALSYSVLSTLTGBEWHGP	64
Qy	62	SAGLNVAAASPYYANMVSVTAGDELTAQVRVAABAYETAYGLTVPPPIAENRAELMTL	121
Db	65	ASASMTAAAAPYYAMVMSVTAVRAEQAGQAERAAAAEAFAFTVPPIVIANRLQLMAL	124
Qy	122	IATNLGONTPALVNEAYGENMAQDAAMFGYAATAATATATLLPFEEPEMTSAGGL	181
Db	125	IATNLVGONPATATAEYQAEWMSODAMMYGYAGASAAAT-QLTPTFPVQTINASGL	183
Qy	182	LEQAAVEEASDTAAA-----NQLMNVPALQLOAQTOGT-----TPPS	222
Db	184	AAQSAATAHATGASAGAQTTLSQLIAAPISVLQGLSSSTAATFASGPGSLGIVGSGS	243
Qy	223	KLGGLMKTVSPHRSPISNMVSMNNHMTNSGVSM-TNTLS---SMLKGFAPAAAAQAV	278
Db	244	WLDKLWALLDPN-----SNFWNTIASSGLFIPSWNTIAPFLGLLGVAADAAGDV	293
Qy	279	QTAQNGVRAMSSLGSSL-----GSSGLGGVVAANIIGRAASYGSLSPVQAWAAANOQVTP	333
Db	294	LGEATSG-----GLGGALVAPLGSAGGLGTGAAGLGNAAATVGTLSVPSPSWTAAAFPLSP	348
Qy	334	AARAL---PLTSLTSAERGPQGMLGPLVGOVGAAGAGGLSGVLRPV----RPVVMPH	386
Db	349	IGSALGTPMVAPPDPAVAG-----NFGMPFGTWGGOGFG-----RAVPQYGRFNFVAR	398
Qy	387	SPAAG 391	
Db	399	PPAAG 403	

RESULT 7

B70931

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70931

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

Db 180 AGQAAATVSTVPPLATTAAVPPQLQLQSLSTSLIPWYSAQQWLAENLLGLTFDNRMTIV 239
QY 225 -----GGLWKTVPSPHSRISPMVSMANNHMTNSGVNTVTLSSMLKGFAPAAA 275
Db 240 RLIGISYFDEGL-----LQFEASIAQQAIPGTPGGAG--DSGSSVLDSWGFTTFA 287
QY 276 QAVQTAQNGVRAMSSL--GSSLSGS-----SGLGGGVAANLGSAAASVGSIS 319
Db 288 -----GPRASPSPVAGGAVGGVQTQPQYWMALDRESIGGSVSAALGXSGSAGSLS 338
QY 320 VPOWMAAANQAVTPAARALP--LTLSTSAARPGQMLGGLPVGQMGARAGGSLGVUR 376
Db 339 VPPDMAARARWANPAARLFGDDVTALRGTAENA---LLRGPFMASAGOSTGGGF--VHK 393
QY 377 VPPRPVMPHSPAAG 391
Db 394 YGFLAVMQRPFPAG 408

RESULT 10

B70625
Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70625
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: B70625
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-391 <COL>
A: Cross-references: GB:292539; GB:AL123456; NID:93261714; PIDN: CAB06873.1; PID: e304546;
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: PPE

Query Match 36.0%; Score 702; DB 2; Length 391;
Best Local Similarity 42.3%; Pred. No. 1.6e-34;
Matches 166; Conservative 62; Mismatches 150; Indels 14; Gaps 7;
QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFOSVVMGLTVGSWIGS 61
Db 1 MDFGALPPEINSARMYAGAGSMAAGAAWNGLAELGTAAASYRSTVITLTTESWMGP 60
QY 62 SAGLMTAAASPYVANKSVTAGQAEELTAQVRVAAAYETAYGLTVPPPPVIAENRAELMTL 121
Db 61 ASMAVMAAQPYLAWITUTYTAEEAAHAGSQAMASAAAAYEAAYATVPPEVVAANRLAAL 120
QY 122 IATNLGQNTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLTPPEAPEMTSAGGL 181
Db 121 VAINVLGINTPATMATEALYAEWKAQDALAYGYAAAG--AAGMLQLPSPSQTNPGGL 179
QY 182 LEQAAVVEASDTAAANQ-----LMNVPAQLQALQOQTGTTTPSKLGLWKTVPSPHS 236
Db 180 AAQSAAVGSAATAAANQVNSVADLISSLPNVAGSLASPTVSLDSTGLSGIITADIALLA 239
QY 237 P--ISNNVMANNHMS--MTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLIG 293
Db 240 TPFVANIINSVNTAAYWYNAIPTALFLANLNSGAPVIAEGATEAAEG---AASAAA 296
QY 294 SSLGSSGLGGVAAANGRAASVGSLSVPQMAAANQVTPAARALPLTSLTSAARPGQ 353
Db 297 AGLADSVTPAGLGSALCEATLVGRLSVPAWSTAAPATTAGATAGLEGSGMTVAABEA-GP 355
QY 354 MLGGLPVGQMGARAGGSLGVLRPPRPYMP 385
Db 356 VTGMMP-GMASAAKGTGAYAGPRYGRKPTVMP 386

RESULT 11

B70560
Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70560
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: B70560
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-413 <COL>
A: Cross-references: GB:295436; GB:AL123456; NID:93261770; PIDN: CAB08826.1; PID: e316565; I
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: PPE

Query Match 36.0%; Score 702; DB 2; Length 413;

Best Local Similarity 39.1%; Pred. No. 1.7e-34;
Matches 168; Conservative 61; Mismatches 145; Indels 56; Gaps 7;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFOSVVMGLTVGSWIG 60
Db 1 MLDFALPPEVNSALMYAGPGSGPMLAAWAAWELQTTASTYDALITGLADGPWQ 60
QY 61 SSAGLMTAAASPYVANKSVTAGQAEELTAQVRVAAAYETAYGLTVPPPPVIAENRAELMI 120
Db 61 SSAASMTAAATPQVWLRSTAGQAEQAGSQAAVAAASAYEAFAFFATVPPEVVAANRLA 120
QY 121 LIATNLGQNTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLTPPEAPEMTSAGG 180
Db 121 LIATNLGQNTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLTPPEAPEMTSAGG 179
QY 181 LLEQAAVVEASDTAAANQLMNVPAQLQALQOQTGTTTPSKLGLWKTVPSPHSRISN 240
Db 180 LASQAASVQAVSGAANAQAALTDIPKAL-----FGLSGIFTNPPPLTLGK 226
QY 241 MVMANNHMTNSGVSMNTLSSMLKGFAPAA---AAQAVQT----- 280
Db 227 ALGLTGHMTSSDGLIVGGVGLDFVQGVTSAGELSDASVAMDTFGKVVSPARLMVTFKD 286
QY 281 -----AAQNGVRAMSSLSGSSLSGSGVGAANLGRASVGSLSVPQA 323
Db 287 YFGLAHDLPKWASEGAKAAGEAKALPAAPVPAIFSAGL-SCVAGAVGQAASVGLKVPV 345
QY 324 WAAANQAVTPAARALPLTSLTSAARPGQMLGGLPVGQMGARAGGSLGVLR--RVPRP 381
Db 346 WTATTPAASPVALAASNLGAAAGSTHAFGGMPL--MGSGAGAFNNPAPRYGPKP 403
QY 382 YMPHSPAAG 391
Db 404 TVIAQPPAGG 413

RESULT 12

C70931
Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: C70931

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-463 <COL>

A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17723.1; PID:e125461

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 35.3%; Score 688.5; DB 2; Length 463;

Best Local Similarity 42.8%; Pred. No. 1.3e-33;

Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;

QY 2 VDFGALPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFGVLPPEINSGRMYAGPGSGPMLAAAWMDGLATELQSTAAADYGSVISVLT-GVWSGQ 59

QY 62 SAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVTAENRAELMIL 121

DB 60 SSGTMAAAAPYVAMSTAAALAREAAQAASAAAAYEAAFAATVPPVVAANRAELAVL 119

QY 122 IATNLGONTPAIAVNEAEYGEWAQDAAMAFGVAATATATATALLPFEAPEMTSAGGL 181

DB 120 AATNIFGONTGAIAAEARYAEMQAQDAAMYGAGSSVAT-CVTFFAAPPTTNAAGL 178

QY 182 LEQAAAVEASDSTAANQLMNNVPOALQOQTAQPTGTTSSKLGGLWKTYS--PHRSP1- 238

DB 179 ATQGVAAQVAGSAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLMAVTVGPVYASSVY 235

QY 239 -----SNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRA 288

DB 236 NSMLGLGFAESKMLVPANDTVISTIFGVQPKFFNPTPNPLIPK----- 283

QY 289 MSSLSGSLG-----SSGLGG--GVAANLGRAASVGLSVFPQMAAANQVTPAARALPL 340

DB 284 -SALGAGLGLRSATSSGLGSTAPALASAGASQAGSVGSGMSPPSWAAATPAIRTVAAVFS 342

QY 341 TSLTS--AAERGPQML-----GGLPVQMGARAGGGLSVLRV 377

DB 343 TGLQAVPAAATSEGLLSQMALASVAGCALGAAARATGGLGGGRV 389

RESULT 13

B70932

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70932

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, M.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70932

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-468 <COL>

A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17730.1; PID:e125462

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 35.0%; Score 682.5; DB 2; Length 468;

Best Local Similarity 40.9%; Pred. No. 2.9e-33;

Matches 164; Conservative 59; Mismatches 125; Indels 53; Gaps 9;

QY 2 VDFGALPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFGLQPPETSGEWMYLGPGAGPMLAAAWMDGLAELQSWAASIASIVEGMASESMLGP 60

QY 62 SAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVTAENRAELMIL 121

DB 61 SSAGMAAAAPYVMTWSGTSACAQAADQAAAVVAYETAFAAVVPQIAARNSQLISL 120

QY 122 IATNLGONTPAIAVNEAEYGEWAQDAAMAFGVAATATATATALLPFEAPEMTSAGGL 181

DB 121 VATNIFGONTAAIAATEAEYGEWAQDTMAEFGYASSATA-SRLTPTTAPPQTTPNSGL 179

QY 182 LEQAAAVEAS-----DTAAANQLMNNVPOALQOQTAQPTGTTSSKLGGL- 227

DB 180 ACQAAATGQATALASGTNAVTTALSSAAAPFPDIIFLLQGLA--TLSTQYTQLMGOLI 237

QY 228 -----WKTYSPIHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQ 276

DB 238 NAIFGPTGATTYQNVFVTAANVTKEFTWANDAMSAPNLGMEFKVF-----WQPPPAPE 291

QY 277 AVQTAQNGVRAMSSLSGLG-----SSGLGGVAAANLGRAASVGLSVPOAWAANQAV 331

DB 292 IPK-----SSLGAGLGLRSAGLSAGLAHAASAGLQANLVGDLSPVPSWASATPAV 341

QY 332 TPAARALPLTSLTA-AERGPQMLGGLPVQMGARAGGGL 371

DB 342 RLVAANTLPATSLAAAPATQIPANLLGQMALGSM---TGCAL 379

RESULT 14

A70646

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: A70646

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, M.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70646

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-380 <COL>

A:Cross-references: GB:Z83867; GB:AL123456; NID:g3261695; PIDN:CAB06278.1; PID:e291015;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match

Best Local Similarity 34.6%; Score 675; DB 2; Length 380;

Matches 172; Conservative 56; Mismatches 143; Indels 28; Gaps 12;

QY 2 VDFGALPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFALLPEVNSARMYTGPGAGSLIAAGGWDLSLAELATTAETAYGSLVSLAALHWRGP 60

QY 62 SAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVTAENRAELMIL 121

DB 61 RAESMAVTAAPYIGWLY-TAEKTOQTALQARAALAFQAVAMTLPPPVVAANRIQLAL 120

QY 122 IATNLGONTPAIAVNEAEYGEWAQDAAMAFGVAATATATALLPFEAPEMTSAGGL 181

DB 121 IATNIFGONTAAIAATEAEYGEWAQDAAMAFGVAATATATALLPFEAPEMTSAGGL 179

QY 182 LEQAAAVEASDSTAANQLMNNVPOALQOQTAQPTGTTSSKLGGLWKTYSPIHRSP1 238

DB 180 TAQAAAVSQATDPLSL--LIETVTCALQALIPSPFIPEDFTFLDAIPAGYATVGTQDVE 237

QY 239 SNMYSM--ANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGL 296

DB 238 SFVACTIGAESNLGLNVNDENPAEVTGDFGIGELVSA----TSPGGGVA-SGAG--- 289

QY 297 GSSGLGGVAAANLGRAASVGLSVPOAWAANQAVTPAARALPLTSL--TSAASERG-PGQ 353

Search completed: June 30, 2004, 16:53:58
Job time : 11.2189 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 6.3659 Seconds
(without alignments)
3198.204 Million cell updates/sec

Title: US-09-597-796c-8

Perfect score: 1949

Sequence: 1 MVDFGALPPEINSARMYAGP.....SGVLRVPRPYVMPHPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1656.5	85.0	396	1 YD61_MYCTU	Q11031 mycobacteri
2	705	36.2	408	1 YS92_MYCTU	Q10813 mycobacteri
3	688.5	35.3	463	1 Y102_MYCTU	O53951 mycobacteri
4	444	22.8	487	1 Y402_MYCTU	P42611 mycobacteri
5	426.5	21.9	443	1 Y878_MYCTU	Q10540 mycobacteri
6	418	21.4	678	1 YF48_MYCTU	Q10778 mycobacteri
7	364.5	18.7	408	1 SRA_MYCLE	Q07297 mycobacteri
8	324.5	16.6	463	1 Y095_MYCTU	Q10892 mycobacteri
9	324	16.6	434	1 YU18_MYCTU	P31500 mycobacteri
10	321.5	16.5	435	1 YU21_MYCTU	O53268 mycobacteri
11	233.5	12.0	178	1 Y129_MYCTU	Q06246 mycobacteri
12	217.5	11.2	176	1 Y125_MYCTU	O50703 mycobacteri
13	153.5	7.9	860	1 ELS_MOUSE	P54320 mus musculu
14	151.5	7.8	232	1 Y125_MYCTU	Q50702 mycobacteri
15	150	7.7	864	1 ELS_RAT	O99372 rattus norv
16	143.5	7.4	730	1 ELS_HUMAN	P15502 homo sapien
17	139.5	7.2	881	1 FR13_YEAST	P47033 saccharomyc
18	135	6.9	907	1 A180_HUMAN	O60641 homo sapien
19	133.5	6.8	825	1 ICP0_HSV2H	P28284 herpes simp
20	133.5	6.8	2432	1 Y43R_IRV6	P18305 chilo iride
21	132	6.8	1120	1 STFR_ECOLI	P76072 escherichia
22	132	6.8	2090	1 N214_HUMAN	P35658 homo sapien
23	131	6.7	836	1 VG26_BPML5	O05233 mycobacteri
24	131	6.7	1150	1 APW0_FIG	P22021 sus scrofa
25	129	6.6	1783	1 RAA3_CHLRE	Q9frec4 chlamydomon
26	128	6.6	779	1 SRP_DROME	P52172 drosophila
27	127.5	6.5	790	1 ANP_NOTOC	P24856 notothenia
28	126	6.5	354	1 VAUG_SCHPO	Q10169 schizosacch
29	125.5	6.4	1211	1 BUN2_DROME	Q24523 drosophila
30	125	6.4	577	1 CST2_HUMAN	P33240 homo sapien
31	124	6.4	2090	1 HFC1_MESAU	P51611 mesocricetu
32	123.5	6.3	677	1 Y136_MYCTU	Q50597 mycobacteri
33	123	6.3	432	1 YF10_MYCTU	P71789 mycobacteri

ALIGNMENTS

RESULT 1

ID	YD61_MYCTU	STANDARD;	PRT;	396 AA.
AC	Q11031;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical PPE-family protein Rv1361c/MT1406.			
GN	Rv1361c OR MT1406 OR MTCY02B10.25C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RX	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Reltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RX	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	J. Bacteriol. 184:5479-5490 (2002).			
CC	!- SIMILARITY: Belongs to the mycobacterial PPE family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; Z75555; CAA99966.1; -			
DR	EMBL; AB007013; AAK45669.1; -			
DR	PIR; H70741; H70741.			
DR	TIGR; W1406; -			
DR	TubercuList; Rv1361c; -			
DR	InterPro; IPR000030; Microbac_PPE.			

34	123	6.3	1140	1 YN96_YEAST	Q04893 saccharomyc
35	122	6.3	394	1 HVF1_ALCEU	P45805 alcaligenes
36	121.5	6.2	1025	1 SLAP_CAUCR	P35828 caulobacter
37	121.5	6.2	1845	1 Z236_HUMAN	O9ul36 homo sapien
38	121	6.2	1199	1 P121_RAT	P52591 rattus norv
39	119.5	6.1	635	1 HMLA_DROME	P10105 drosophila
40	119	6.1	915	1 A180_RAT	Q05140 rattus norv
41	118.5	6.1	580	1 EXPR_XANCP	P23314 xanthomonas
42	118.5	6.1	2038	1 FSH_DROME	P13709 drosophila
43	118	6.1	1508	1 BCSG_XANAC	P58938 xanthomonas
44	117	6.0	444	1 Y808_CHLPN	Q92798 chlamydia p
45	117	6.0	652	1 PICA_HUMAN	Q13492 homo sapien

```

DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 158 159 TA -> AT (IN REF. 2).
SQ SEQUENCE 396 AA; 40015 MW; 6AFAE0D7B5F668D0 CRC64;

Query Match
Best Local Similarity 85.0%; Score 1656.5; DB 1; Length 396;
Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMVAASPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
DB 61 SSAGLMVAASPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLTPPEEAPSWTSAGG 180
DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATLTPPEEAPSWTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVQALQQAQPTGTPSSKGLGKWTVPSPHRSPI 240
DB 181 LLEQAAVEEADTAAANQLMNNVQALQQAQPTKSIWPPDQLSELKWAISPHLSPLN 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVCTAAQNGVRAMSS----LGSSL 296
DB 241 IVSMNNHMTNSGVSMNTLSSMLKGFAPAAAQAQVCTAAQNGVRAMSSLGSLGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGLSVFPQAWAANQAVTPAARALPLTSLTSAERPGQMLG 356
DB 300 GSSGLGAGVAAANLGRAASVGLSVFPQAWAANQAVTPAARALPLTSLTSAQAPGHMLG 359
QY 357 GLPVQOMCARAG--GGLSGVLRVPRPYVMPHSPAAG 391
DB 360 GLPLQLGTLNSGGGFGVSNALRMPPRAYVMPRVPAAG 396

RESULT 2
Y92 MYCTU STANDARD; PRT; 408 AA.
AC Q10813;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Hypothetical PPE-family protein RV2892c/MT2959/MB2916c.
GN RV2892C OR MT2959 OR MICY274-23C OR MB2916C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544 (1998).
[2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

```

```

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -/- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; Z74024; CAA98377.1; -.
DR ENBL; AE007119; AAK47285.1; -.
DR ENBL; BX248344; CAD96603.1; -.
DR PIR; G70925; G70925.
DR TIGR; MT2959; -.
DR Tuberculist; RV2892C; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 56 76 POTENTIAL.
SQ SEQUENCE 408 AA; 41469 MW; 3E3D1F20D7827199 CRC64;

Query Match
Best Local Similarity 36.2%; Score 705; DB 1; Length 408;
Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIGS 61
DB 1 MDGVLPEINSGRMYAGPGSGPMWMAAANDSLAELGLAAGGVRLAISLTGAYWAGP 60
QY 62 SAGLMVAASPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVPIAENRAELMI 121
DB 61 AAASVVAATPVVAVLSATAGCAQGAQQAARAAAAYELAFAMTVPVPPVVAANLALLVAL 120
QY 122 IATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLTPPEEAPSWTSAGGL 181
DB 121 VATNFFGQNTPAIAATEAQVAAEMWQAQDAAMVAVAGSAIAT-ELTPTTAPVTTSPAAL 179
QY 182 LEQAAA-----VEEASDTAAANQLMNN-----VP--QAQO--LAQPTGQTPPSKGL--- 224
DB 180 AQCAAAATSVSTVPPLATTAAPVQLLQSLSTSLIPWTSALQWLAEALLGLTPDNRTIV 239
QY 235 -----GGLWKTVPSPHRSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAA 275
DB 240 RLGLSYDFEGL-----LQFEASLAQAQAPGTPGGAG--DSGSSVLDSNGPTIFA 287
QY 276 QAVQTAQNGVRAMSSL--GSSLGS-----SGLGGVAAANLGRAASVGLSLS 319
DB 288 -----GPRASPVAGGGAUGVQTPQPYWYWDRESIGGSVGAALKGSSAGLSLS 338
QY 320 VPQAAANQAVTPAARALP---LTSLSAERPGQMLGGLPVQOMCARAGGGLSGVLR 376
DB 339 VFPDWAARWANFAAWRLPGDDVTALRGTAENA---LLRGFPNWSAGQSTGGGF--VHK 393
QY 377 VPPRPVYVMPHSPAAG 391
DB 394 YGFRLLAVQVRPFFAG 408

```

```

RESULT 3
Y102 MYCTU
ID Y102 MYCTU STANDARD; PRT; 463 AA.
AC 053951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1802/MT1851/MB1830.
GN Rv1802 OR MT1851 OR MT049-24 OR MB1830.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Barrell B.G.;
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Allard D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsepe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -1- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AL022021; CAA17723.1; -
CC ENBL; AF007044; AAK46123.1; -
CC ENBL; BX248340; CAD94533.1; -
CC PIR; C70931; C70931.
CC TIGR; MT1851; -
CC TubercuList; Rv1802; -
CC InterPro; IPR000030; Microbac_ppp.
CC Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 401 401 S -> L (IN REF. 2).

```

```

SQ SEQUENCE 463 AA; 46021 MW; EE64828BF09FA551 CRC64;
Query Match 35.3%; Score 688.5; DB 1; Length 463;
Best Local Similarity 42.8%; Pred. No. 2.8e-33;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;
QY 2 VDFGALPPEINSARMYAGPGSASLVRAAQWMDSVASDLFSAAGAFQSVVWGLTVGWSIGS 61
DB 1 MDFGVLPPEINSGRMYAGPGSGPMLAAAAWDGLATELQSTADYGVSVILT-GWWSGQ 59
QY 62 SAGLWVAAGPYVAWMSVTAGQAELETAQVRVRAAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 60 SSGTMAAAAAAPYVAWMSATLAALAEAAQAASAAAAAYEAAFAATVTPPPVVAANRAELAVL 119
QY 122 IATNLLGQNTPAIIVNEAEVGEWMAQDAAMFGVAAATATATATALLBEEBEPMTSAGGL 181
DB 120 AATNIFGQNTGAIAAEAEARVMAQDAAMYGAGSSVAT-QVTFFAAPPTTNAAGL 178
QY 182 LEQAAVEEASDTAAANQNMNVPOALQQLAQPTQGTTPSSKLGGLWKTVS--PHRSP1- 238
DB 179 ATQGVAAVQAQVAGASAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLNAVTVGVYASSVY 235
QY 239 -----SNMYSMANNNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAQVTAQAQNGVRA 288
DB 236 NSMLGLGFAESKMWLPANDTVISTIFGMVOPKQFFNPVTFNPDLIPK----- 283
QY 289 MSSLGSSIG-----SSGLGG--GVAANLGRPAASVGLSVPAQAAANQAVTPAARALEL 340
DB 284 -SALGAGLGLRSAISSGLGSTAPPAISAGASQAGSVGMSVPPSWAAATPAIRTVAANFSS 342
QY 341 TSLTS--AAERGPQQL-----GGLPVQMGARAGGGLSGVLRV 377
DB 343 TGLQAVFAAAISEGSLLSQMALASVAGGALGCAAAARATGGFLGGRV 389
RESULT 4
Y442 MYCTU
ID Y442 MYCTU STANDARD; PRT; 487 AA.
AC P42611; O53727;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv0442c/MT0458.
GN Rv0442C OR MT0458 OR MT037.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Erdmann;
RX MEDLINE=87137260; PubMed=3029018;
RA Shinnick T.M.;
RT "The 65-kilodalton antigen of Mycobacterium tuberculosis.";
RL J. Bacteriol. 169:1080-1088 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Barrell B.G.;
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

```


[illegible]

DR	InterPro: IPR000030; Microbac_PPE.	
DR	Pfam: PF00823; PPE; 1.	
KW	Hypothetical protein; Transmembrane; Complete proteome.	
FT	TRANSMEM 3 23	POTENTIAL.
FT	TRANSMEM 88 108	POTENTIAL.
FT	TRANSMEM 112 132	POTENTIAL.
FT	TRANSMEM 216 236	POTENTIAL.
FT	TRANSMEM 245 265	POTENTIAL.
FT	TRANSMEM 275 296	POTENTIAL.
FT	TRANSMEM 323 343	POTENTIAL.
FT	TRANSMEM 419 439	POTENTIAL.
SQ	SEQUENCE 463 AA; 46894 MW; 42D9D66A033D0DD8 CRC64;	
Query Match 16.6%; Score 324.5; DB 1; Length 463;		
Best Local Similarity 27.8%; Pred. No. 4e-12;		
Matches 112; Conservative 54; Mismatches 144; Indels 93; Gaps 12;		
QY	6 ALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGWSGAGL 65	
DB	2 AIPPEVHSGLLSAGCGFGLLVAAQWQELSQQYALACAEQLQGLGEVQASWQGTAAQ 61	
QY	66 MVAASGPYVAMVMTAGQAEELTAQVRVAAAYETAYGLTVPPVIAENRAELMILIATN 125	
DB	62 YVAHGFYLANLEQTAINSVTAQHVAHAAAYCSALAAMPITPAELAAHNAHGVLIATN 121	
QY	126 LLQONTPTAIVNBAEYGEHWAQDAAMFGVAAATATATATLLPPEAPDWTAGGLLEQA 185	
DB	122 FFGINTVPIALNEADYVRWMLQADTMAAYQAVADAATVAVPSTQPPAPPIRAPGG 176	
QY	186 AAVEEASDT-----AAANOLMNVVQALQOLAQPTQ-----GTPSSKL----- 224	
DB	177 -----DAADTDLVSSIGQIIRI-----LDPIAPYKTFLEFFQFGFSAVTVVLALVAL 229	
QY	225 ---GGLW---KTVSPHRSPISNVMVMANNHMTNSGVSMTNTLSSMLK-----GF 269	
DB	230 QLYDFLWYPPYAYGYGLLLFFFTP-----TLTALTALSAHLINLPPAGL 275	
QY	270 APAAQAQVQTAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVQPAWAAANQ 329	
DB	276 LPFAAA-----LGPQDQGANLAVAVTATAAVP-----GGSP 308	
QY	330 AVTPAPARALPLTSLTSAERGGP---QMLGGLPVG-QMGARAG 368	
DB	309 PTNPAPAAASNSVGSASAPGISVAVGLAPPGVSSGPKAG 351	
RESULT 9		
YU18	MYCTU	
ID	YU18_MYCTU	
AC	P31500; O53265; STANDARD; PRT; 434 AA.	
DT	01-JUN-1993 (Rel. 26, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	10-OCT-2003 (Rel. 42, Last annotation update)	
DE	Hypothetical PPE-family protein Rv3018c/MT3098/MT3101.	
GN	Rv3018c OR MT3098/MT3101 OR MTV012.32C.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1773;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H37Rv;	
EX	MEDLINE=98295987; PubMed=9634230;	
RA	Cole S.T., Brosch K., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.,	
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,	
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,	
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,	
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,	
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;	
RT	"Deciphering the biology of Mycobacterium tuberculosis from the	
RT	complete genome sequence."	

RL	Nature 393:537-544 (1998).	
[2]		
RN	SEQUENCE FROM N.A.	
RP	STRAIN=CDC 1551 / Oshkosh;	
RC	MEDLINE=22208494; PubMed=12218036;	
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,	
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,	
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,	
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,	
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;	
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and	
RT	laboratory strains";	
RL	J. Bacteriol. 184:5479-5490 (2002).	
[3]		
RN	SEQUENCE OF 160-374 FROM N.A.	
RP	STRAIN=Isolate 50410;	
RC	Patki A.H., Dale J.W.;	
EL	Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.	
CC	-!- SIMILARITY: Belongs to the mycobacterial PPE family.	
CC	-!- CAUTION: In strain Oshkosh the gene for this protein is	
CC	interrupted in position 307 by an 18610 element.	
CC	-!- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate	
CC	reductase.	
CC	-!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts	
CC	in positions 294; 337 and 355.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AL021287; CA416103.1; -	
DR	EMBL; AE007129; AAK47427.1; ALT_SEQ.	
DR	EMBL; AE007129; AAK47430.1; ALT_SEQ.	
DR	EMBL; X59271; CAA41961.1; ALT_FRAME.	
DR	PIR; E70857; E70857.	
DR	TIGR; MT3098; -	
DR	TIGR; MT3101; -	
DR	TubercuList; RV3018c; -	
DR	InterPro; IPR000030; Microbac_PPE.	
DR	Pfam; PF00823; PPE; 1.	
KW	Hypothetical protein; Complete proteome.	
SQ	SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;	
Query Match 16.6%; Score 324; DB 1; Length 434;		
Best Local Similarity 28.0%; Pred. No. 4e-12;		
Matches 115; Conservative 56; Mismatches 178; Indels 62; Gaps 12;		
QY	6 ALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGWSGAGL 65	
DB	8 ASPPEVHSGALLSAGPGSLQAAAGNSALSAEYAAVQELSUVVAAVAGVWQGPSAEL 67	
QY	66 MVAASGPYVAMVMTAGQAEELTAQVRVAAAYETAYGLTVPPVIAENRAELMILIATN 125	
DB	68 FVAAYVPVAVLVQASADSAAAAGHEHAAAGVYCALAEMPTLPBLAAHNLTHAVLVATN 127	
QY	126 LLQONTPTAIVNBAEYGEHWAQDAAMFGVAAATATATATLLPPEAPDWTAGGLLEQA 185	
DB	128 FFGINTVPIALNEADYVRWVWQAATVNGAYEAVVGAALVAVPHTGPAPVIVKPG 181	
QY	186 AAVEEASDTAAAN-----QLMNVVQALQOLAQPTQGTTPSSKLGGLKWTVSFPH 234	
DB	182 --ANEASNAVAATITPPPWHEIVQLEETFAAYQYLSALLSELPA--VAVWVQFLFVD 237	
QY	235 ---RSPISNMVSMANNHMTNSGVSMTNTLSSMLKGA----- 270	
DB	238 ILGFNIIGIITILASNAQILLTEFAINASYVAVGLLYATAGVIDIVVWVIGNLFGVVPIL 297	
QY	271 -----PRAAQAVQTAQNGVYRAMSSLGSSGLGGVAAANLGRAASV--GSLSVPOA 323	

RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z95389; CAB08678.1; -;
DR EMBL; AEO07158; AKA47873.1; -;
DR PIR; C70975; C70975.
DR TIGR; MT3533; -;
DR TubercuList; RV3429; -;
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 178 AA; 19811 MW; 8BE1FC025ABFBEA6 CRC64;

Query Match 12.0%; Score 233.5; DB 1; Length 178;
Best Local Similarity 35.9%; Pred. No. 2.8e-07;
Matches 65; Conservative 25; Mismatches 76; Indels 15; Gaps 3;

QY 7 LPPEINARMYAGPGSASLVAAQ---MWDS---VASDLFSAASAFQSVVWGLTVGSMI 59
DB 5 IPAEYISNIIYEGPGADSLFASQGLRELAYSVETTAESLEDELD-ENWKGSSDIL 56
QY 60 GSSAGLMAVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELM 119
DB 57 GSSSDLMAADAGYLDWLTGHSQILETAVVIDFLAVVYETTRHKVVPATIANNEEVH 116
QY 120 ILIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPMTSAG 179
DB 117 RLIAASNAGVNTPAIAGLDAQYQYRAQNTAVMNDYQSTARFILAYLPRWQEPPIYGGG 176
QY 180 G 180
DB 177 G 177

RESULT 12
ID_VY25_MYCTU STANDARD; PRT; 176 AA.
AC Q50703;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV3425.
GN RV3425 OR MTCY78.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tebala R.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544(1998).

CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z77165; CAB01031.1; -;
DR PIR; F70738; F70738.
DR TubercuList; RV3425; -;
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 176 AA; 19855 MW; B8CFE2E9463B87B0 CRC64;

Query Match 11.2%; Score 217.5; DB 1; Length 176;
Best Local Similarity 33.5%; Pred. No. 2.3e-06;
Matches 58; Conservative 31; Mismatches 83; Indels 1; Gaps 1;

QY 7 LPPEINARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSGSAGLM 66
DB 5 IPAEYISNIIYEGPGADSLFASQGLRELAYSVETTAESLEDELD-ENWKGSSDIL 63
QY 67 VAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMILITNL 126
DB 64 ADAVERYLQWLSKSSQLKHAAWINGLANAYNDRKRWPPPEIAANREERRRLIASNV 123
QY 127 LGONTPATVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPMTSAG 179
DB 124 AGVNTPAIADLDAQYDQVRNVAVMNAVSVTSLSLDLPWREPPPIYGGG 176
QY 127 LGONTPATVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPMTSAG 179
DB 124 AGVNTPAIADLDAQYDQVRNVAVMNAVSVTSLSLDLPWREPPPIYGGG 176
QY 127 LGONTPATVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPMTSAG 179
DB 124 AGVNTPAIADLDAQYDQVRNVAVMNAVSVTSLSLDLPWREPPPIYGGG 176

RESULT 13
ID_ELS_MOUSE STANDARD; PRT; 860 AA.
AC F54320;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Elastin precursor (Tropoelastin).
GN ELN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Lung;
RX MEDLINE=95133069; PubMed=7429060;
RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT "Use of an intron polymorphism to localize the tropoelastin gene to
RT mouse chromosome 5 in a region of linkage conservation with human
RT chromosome 7.";
RL Genomics 23:125-131(1994).
CC -!- FUNCTION: Major structural protein of tissues such as aorta and
CC nuchal ligament, which must expand rapidly and recover completely.
CC -!- SUBUNIT: The polymeric elastin chains are cross-linked together
CC into an extensible 3D network.
CC -!- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.
CC -!- PTM: The crosslinks are made of deaminated Lys.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U08210; AAA80155.1; -;

CC -!- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=8;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;

CC Name=1;
 CC IsoId=Q99372-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q99372-2; Sequence=VSP_004244;
 CC Name=3;
 CC IsoId=Q99372-3; Sequence=VSP_004245;
 CC Name=4;
 CC IsoId=Q99372-4; Sequence=VSP_004246;
 CC Name=5;
 CC IsoId=Q99372-5; Sequence=VSP_004244, VSP_004245;
 CC Name=6;
 CC IsoId=Q99372-6; Sequence=VSP_004245, VSP_004246;
 CC Name=7;
 CC IsoId=Q99372-7; Sequence=VSP_004244, VSP_004246;
 CC Name=8;
 CC IsoId=Q99372-8; Sequence=VSP_004244, VSP_004245, VSP_004246;
 CC -!- PTM: The crosslinks are made of deaminated Lys.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M60647; AAA42269.1; -;
 DR EMBL; J04035; AAA42268.1; -;
 DR EMBL; M86372; AAA42271.1; -;
 DR EMBL; M86355; AAA42271.1; JOINED.
 DR EMBL; M86363; AAA42271.1; JOINED.
 DR EMBL; M86364; AAA42271.1; JOINED.
 DR EMBL; M86366; AAA42271.1; JOINED.
 DR EMBL; M86371; AAA42271.1; JOINED.
 DR EMBL; M86376; AAA42272.1; -;
 DR EMBL; M86373; AAA42272.1; JOINED.
 DR EMBL; M86375; AAA42272.1; JOINED.
 DR PIR; A36106; EART.
 DR InterPro; IPR003979; tropoelastin.
 DR PRINTS; PR01500; TROPOELASTIN.
 DR Structural protein; Connective tissue; Repeat; Signal;
 KW Alternative splicing.
 FT NON TER 1
 FT SIGNAL <1 21
 FT CHAIN 22 864
 FT DISULFID 854 859
 FT VARSPPLIC 263 307
 FT BY SIMILARITY.
 FT ELASTIN.
 FT BY SIMILARITY.
 FT Missing (in isoform 2, isoform 5, isoform
 FT 7 and isoform 8).
 FT /FtId=VSP_004244.
 FT Missing (in isoform 3, isoform 5, isoform
 FT 6 and isoform 8).
 FT /FtId=VSP_004245.
 FT Missing (in isoform 4, isoform 6, isoform
 FT 7 and isoform 8).
 FT /FtId=VSP_004246.
 FT SEQUENCE 864 AA; 72786 MW; 456894BB09E79FD4 CRC64;

Query Match 7.7%; Score 150; DB 1; Length 864;
 Best Local Similarity 25.9%; Pred. No. 0.11;
 Matches 102; Conservative 31; Mismatches 153; Indels 108; Gaps 18;

Qy 5 GALT-----PEINSMYAGPGSGLVAAQWDSVADLFSASAFOSVWGLTVGSW-- 58
 Db 399 GGIPGVGGPGIGGPGIVGGPGCAVPAKAAKAAK--YGARGGVGPIYGVGAGGPG 456
 Qy 59 --IGSSAGLMVAASPVMNMTAGOAELTAQVRVAAAYETAYGLTVPPIAENRA 116
 Db 457 YGVGAGAGL-----GGASQAAA-----AAAKAKYAG----- 486

Qy 117 ELMILIIATNLGONTPAIVNEABYGEWMAODAAAMFGYAAATATATATLLPFEAEPM 176
 Db 487 -----GAGTLGGLVPG-----AVPGALPGAVPGALPGAVPGALP-GAVEGVP 527
 Qy 177 SAGGLL---EQAAAVEBASDTAAANQLMNNVPOALQLOPTQGTTFSS-KLGGLWKTVS 232
 Db 528 GTGGVPGAGTTPAAAAAATAKAAA-----KAGQYGLGPGVGGVGGVGGVGGVGG 579
 Qy 233 PHRSPISNMVSMMNNHSMSTNSGVSM-TNTLSSMLKGF-APAAAAAQAVQTAAONGVTRMS 290
 Db 580 FG-----GVTGIGTGPGLVPGDLGGAGTPAAAKSAKAAKAAKAAKAYRAAA 624
 Qy 291 SLGSSLGSSGLGGV-----AANLGRAASV---GSLSVPCAAMAAANQAVTPPARALPLT 341
 Db 625 GLGAGVPLGVGAGVPGFAGAGGFGAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGF 374
 Qy 342 SLTSAAEERPGQMLGGLPVGQMGARAG-GGLSGV 708
 Db 682 -----GGPGGLGGPGGLGGPGGGGGGGLGGV 708

Search completed: June 30, 2004, 16:49:44
 Job time : 9.3659 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 28.479 Seconds
(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796c-8
Perfect score: 1949
Sequence: 1 WDFGALPEINSARYAGP.....SGVLRVPRPYMPHPSPAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phase:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virine:
16: sp_bacteriaph:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	100.0	391	16	O05298 mycobacteri
2	1920.5	98.5	330	16	Q7U0E9 mycobacteri
3	1779.5	91.3	336	16	Q7U071 mycobacteri
4	1648	84.6	393	16	O8VI23 mycobacteri
5	1583	81.2	393	16	O06341 mycobacteri
6	1583	81.2	393	16	Q7W4F5 mycobacteri
7	779.5	40.0	393	16	Q7T2J3 mycobacteri
8	779.5	40.0	411	16	O53939 mycobacteri
9	754	38.7	410	2	Q99Q11 mycobacteri
10	751	38.5	409	16	O53957 mycobacteri
11	749	38.4	409	16	Q7T2H7 mycobacteri
12	740.5	38.0	399	16	Q7T2H8 mycobacteri
13	737.5	37.8	403	16	O53956 mycobacteri
14	736	37.8	423	16	O53950 mycobacteri
15	736	37.8	423	16	Q7T2I4 mycobacteri
16	731	37.5	421	16	Q9Z5K0 mycobacteri

17	702	36.0	391	16	P96362 mycobacteri
18	702	36.0	391	16	Q7U0T5 mycobacteri
19	702	36.0	413	16	O06386 mycobacteri
20	683.5	35.1	694	16	Q8VUW0 mycobacteri
21	682.5	35.0	468	16	O53958 mycobacteri
22	675	34.6	380	16	P95190 mycobacteri
23	675	34.6	380	16	Q7TX66 mycobacteri
24	670.5	34.4	385	16	Q7T2R7 mycobacteri
25	669.5	34.4	394	16	Q7TXX5 mycobacteri
26	668.5	34.3	462	16	O33310 mycobacteri
27	667.5	34.2	385	16	O33204 mycobacteri
28	666.5	34.2	385	16	Q8VU20 mycobacteri
29	641	32.9	364	16	Q7T2J5 mycobacteri
30	637	32.7	350	16	Q7T2J2 mycobacteri
31	635	32.6	363	16	O53940 mycobacteri
32	633.5	32.5	365	16	O86373 mycobacteri
33	629.5	32.3	405	16	Q8VUW5 mycobacteri
34	626.5	32.1	381	16	Q7TX67 mycobacteri
35	618.5	31.7	397	2	Q9AGF0 mycobacteri
36	615.5	31.6	382	16	Q7TXX3 mycobacteri
37	615.5	31.6	402	16	O33312 mycobacteri
38	610	31.3	443	16	O8VKL9 mycobacteri
39	604	31.0	443	16	Q7U242 mycobacteri
40	603	30.9	426	16	O05907 mycobacteri
41	603	30.9	391	16	O05798 mycobacteri
42	597.5	30.7	391	16	Q7TX76 mycobacteri
43	594.5	30.5	391	16	Q7TX76 mycobacteri
44	589	30.2	406	16	P71869 mycobacteri
45	589	30.2	406	16	Q7TW99 mycobacteri

ALIGNMENTS

RESULT 1

O05298 PRELIMINARY; PRT; 391 AA.
ID O05298
AC O05298;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein (PPE family protein).
GN RV1196 OR MTCU364.08 OR MTL234.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv.
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Ohkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.


```

QY 181 LLEQAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHRSPI 240
DB 181 LLEQAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHLSPI 240
QY 241 MYSMANNHMTNMGVSMNTTSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGSSL 296
DB 241 IVSMLNHNHMTNMGVSMNTTSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAARPGQM 356
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPPAWAANQAVTPAARALPLTSLTSAAG 359
QY 357 GLPVGOMGARAG--GGLSGVLVPPRPVPHSPHSPAG 391
DB 360 GLPLGOLTNSSGGGFGVSNALRMPPRAYVMPVPAAG 396

RESULT 4
Q8VIZ3
ID Q8VIZ3 PRELIMINARY; PRT; 393 AA.
AC Q8VIZ3;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007161; AAK47941.1; -.
DR TIGR; MT3582; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR000508; Peptidase_S26.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE I.1; 1.
SQ SEQUENCE 393 AA; 39688 MW; 86F0B67798855511 CRC64;

Query Match 84.8%; Score 1648; DB 16; Length 393;
Best Local Similarity 84.8%; Pred. No. 2e-85;
Matches 395; Conservative 20; Mismatches 34; Indels 6; Gaps 3;

QY 1 MVDFGALPPINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSM 60
DB 1 MVDFGALPPINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSM 60
QY 61 SSAGLMVAASPYVAMSVTAGAETAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASPYVAMSVTAGAETAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTPAIVNNEAYGEWQAQDAAMFGYAAATATATATALLPPEEAPMTSAG 180
DB 121 LIATNLGONTPAIVNNEAYGEWQAQDAAMFGYAAATATATATALLPPEEAPMTSAG 180
QY 181 LLEQAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHRSPI 240
DB 181 LLEQAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHLSPI 240
QY 241 MYSMANNHMTNMGVSMNTTSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGSSL 296
DB 241 LLEQAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHLSPI 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAARPGQM 356

```

```

DB 241 IVSMLNHNHMTNMGVSMNTTSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAARPGQM 356
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPPAWAANQAVTPAARALPLTSLTSAAG 359
QY 357 GLPVGOMGARAG--GGLSGVLVPPRPVPHSPHSPAG 391
DB 360 GLPLGOLTNSSGGGFGVSNALRMPPRAYVMPVPAAG 396

RESULT 5
Q06341
ID Q06341 PRELIMINARY; PRT; 393 AA.
AC Q06341;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR MTCY113E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RL MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
DR EMBL; Z95390; CAB08702.1; -.
DR PIR; C70568; C70568.
DR Tuberculist; RV3478; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein: Complete proteome.
KW SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 81.2%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 9.3e-82;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDFGALPPINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSM 60
DB 1 MVDFGALPPINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSM 60
QY 61 SSAGLMVAASPYVAMSVTAGAETAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASPYVAMSVTAGAETAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTPAIVNNEAYGEWQAQDAAMFGYAAATATATATALLPPEEAPMTSAG 180
DB 121 LIATNLGONTPAIVNNEAYGEWQAQDAAMFGYAAATATATATALLPPEEAPMTSAG 180
QY 181 LLEQAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHRSPI 240
DB 181 LLEQAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHLSPI 240
QY 241 MYSMANNHMTNMGVSMNTTSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGSSL 296
DB 241 VSSIANNHMTNMGVSMNTTSSMLKGLAPAAAQAQVTAONGVRAMSS-----LGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAARPGQM 356

```

```

Db 300 GSSGLGAGVAANLGRAASVGLSVPPAWAANQAVTPAARALPLTSLTSAATAPGHMLG 359
QY 357 GLPVGCMGARAGGLSVGLRVPPRPVPMHPSAAG 391
Db 360 GLPLGH-SVNGSGINNLRVPARAYAIPTPTPAAG 393

RESULT 6
Q7TWFS PRELIMINARY; PRT; 393 AA.
AC Q7TWFS;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE PE family protein.
CN PP260 OR ME3505.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RX MEDLINE=22709107;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248346; CAD95692.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF420C95DAE7DD4 CRC64;

Query Match 81.2%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 9.3e-82;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Db 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLWAAASPYVAMSVTAGQAEELTAQVRVAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWAAASPYVAMSVTAGQAEELTAQVRVAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
QY 181 LLEQAAVVEEADTAAANQLMNNVPOALQLOAQPTQCTTSSKLGGLWKTVPSPIN 240
Db 181 LLEQAAVVEEADTAAANQLMNNVPOALQLOAQPTQCTTSSKLGGLWKTVPSPIN 240
QY 241 MVSMANNHMTSGVNTLTSSMLKGFAPAAAQVQAAQVQAAQVQAAQVQAAQVQAAQV 296
Db 241 VSSANNHMTSGVNTLTSSMLKGFAPAAAQVQAAQVQAAQVQAAQVQAAQVQAAQV 299
QY 297 GSSGLGGVVAANLGRAASVGLSVPPAWAANQAVTPAARALPLTSLTSAAGRGPMQLG 356
Db 297 GSSGLGGVVAANLGRAASVGLSVPPAWAANQAVTPAARALPLTSLTSAAGRGPMQLG 359
QY 357 GLPVGCMGARAGGLSVGLRVPPRPVPMHPSAAG 391
Db 360 GLPLGH-SVNGSGINNLRVPARAYAIPTPTPAAG 393

RESULT 7
Q7TWJ3 PRELIMINARY; PRT; 393 AA.
AC Q7TWJ3;
DT 01-OCT-2003 (TReMBLrel. 25, Created)

```

```

DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE PPE family protein.
CN PPE26 OR ME1817.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248340; CAD94520.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 38572 MW; 41944B6E547A8AB0 CRC64;

Query Match 40.0%; Score 779.5; DB 16; Length 393;
Best Local Similarity 43.7%; Pred. No. 2e-36;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIGS 61
Db 1 MDGFGALPPEVNSVRMYAGPGSAPMVAAASAWNGLAELSSAATGVETITQLSSEGMLGP 60
QY 62 SAGLWAAASPYVAMSVTAGQAEELTAQVRVAAAYETAYGLTVPPVIAENRAELMI 121
Db 61 ASANAAMAEVAPYVAMSVTAQAQAEQATQARAAAFERAAFAATVPPPLIANNASLMQL 120
QY 122 IATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGGL 181
Db 121 ISTNVFGQNTSAIAAAEAQYGEWMAQDSAAVMYAVAGSSASASA-VTPFSTPPQIANPTAQ 179
QY 182 LEQAAVVEEADTAAANQLMNNVPOALQLOAQPTQCTTSSKLGGLWKTVPSPIN 229
Db 180 GTQAAVATAAGTAACTTTEMITGLPNALQSLTLPQLQSS-NGPLSLMLWQLFOTPNPPT 238
QY 230 -----TVSPHRSPISNNVSMANNHMTSGVNTLTSSMLKGFAPAAAQVQAAQVQAA 282
Db 239 SISALLTDLQYAFSEFFNTGELPYFSGNGNFFQAAKTL-GLIGSAAPAAVA-----AA 292
QY 293 QNGVAMSSGLSSGLGGVVAANLGRAASVGLSVPPAWAANQAVTPAARALPLT 341
Db 293 GDAAGLPLGLGMLG----GGPVAAGLGNAAASVGLSVPPVWMSGPLGSGVTFGAAPLPVS 348
QY 342 SLTSAAGRGPMQLGGLPVGQMGAPAGGLSVGLRVPPRPVPMHPSAAG 391
Db 349 TVSAAPAEAPGSLGGLPL-----AGAGGAGAP-RYGRPTVMARPPFAG 393

RESULT 8
O53939 PRELIMINARY; PRT; 411 AA.
AC O53939;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE PPE family protein.
GN RV1789 OR MT1838 OR MTV049.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

```

```
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL020201; CAAL7711.1; ALT_INIT.
DR EMBL; AF007043; AAK46108.1; -.
DR PIR; G70929; G70929.
DR TIGR; MT1838; -.
DR TubercuList; Rv1789; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E059087DA CRC64;

Query Match 40.0%; Score 779.5; DB 16; Length 411;
Best Local Similarity 43.7%; Pred. No. 2.1e-36;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VFPGALPPPEINSGARMYVAGPGSASLVAAQWDSVSDLSFAASAFQSVVWGLTVGSWIGS 61
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 19 MDFGALPPPEINSGARMYVAGPGSAPVMAAGAWGLAAELSSAATGVTETVITQLSS 78
QY 62 SAGLMVAASPYVWMSVTTAGQAELETAQVRAAAAYETAYGLTVPPPPVIAENRAELMI 121
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 79 ASNMAEAVPYVWMSAAQAQAEQATQARAAAAPFAATVPPPLIANNRASLMQL 138
QY 122 IATNLGONTPAIVNAEYEGEMWADAAAFYGAATATATATATLPPFEAPEMTSAGL 181
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 139 ISTNVFGQNTSAIAAAEAQYEGEMWADAAAFYGAATATATATLPPFEAPEMTSAGL 197
QY 182 LEQAAAVEASDTAAA--NQLMNVVPAALQALQPTGCTTPSSKGLGLWK----- 229
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 198 GTQAAAVATAAGTAQSTLTEMITGLPNALQSLTSLPQLSS--NGPLSWLMQILFCTPN 256
QY 230 -----TVSPHRSPISNMVMANNHMSMTNSGV---SMTNTLSSMLKGFAPAAAQAVQTA 282
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 257 SISALLTDIQPYASFFYNTGELPYFSGIGNNFIQSAKTL-GLIGSAPAAVA-----AA 310
QY 283 QNGVRAVSSIGSLGSGGLGGVAAVNLGRAASVGSLSVPOAWAA--ANQAVTPAARALPLT 341
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 311 GDAAGKLPGLGGLMG---GGPVAAGLGNAAVSGKLSVPPVMSGSLPFGSVTFCGAAPL 366
QY 342 SLTSAERGGOMLGLPVQMGARAGGLSGVLRVPPVYPVPHSPAAG 391
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 367 TVSAAPAPGSLGGLPL-----AGAGGAGAGP-RYGFRPTVWVAPFPAG 411

RESULT 9
Q99Q11 ID Q99Q11 PRELIMINARY; PRT; 410 AA.
AC Q99Q11; 2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Rv1808-like protein.
GN MYC1808 OR OV1808.
```

```
OS Mycobacterium microti.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1806;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MYC 94-2272, and OV254;
RA Sivadon V., Heym B., Mazancourt P., Gaillard J.-L.L.;
RT "PPE Rv1808 orthologue of Mycobacterium microti."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF335180; AAK20894.1; -.
DR EMBL; AF335179; AAK20893.1; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR SEQUENCE 410 AA; 40044 MW; 3B9DA3174655A5EA CRC64;

Query Match 38.7%; Score 754; DB 2; Length 410;
Best Local Similarity 44.5%; Pred. No. 5.7e-35;
Matches 177; Conservative 51; Mismatches 140; Indels 30; Gaps 10;

QY 1 MVDFGALPPPEINSGARMYVAGPGSASLVAAQWDSVSDLSFAASAFQSVVWGLTVGSWIG 60
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 1 MLDFGALPPPEINSGARMYVAGPGSGPELLAAANADALAAELYSAAASYGTIEGLTVAPWM 60
QY 61 SSAGLMVAASPYVWMSVTTAGQAELETAQVRAAAAYETAYGLTVPPPPVIAENRAELMI 120
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 61 PSSITMAAAVAPYVWMSVTTAGQAEQAGAKIAAGYVETAFATVPPPPVIEANRALLMS 120
QY 121 LIATNLGONTPAIVNAEYEGEMWADAAAFYGAATATATATLPPFEAPEMTSAGG 180
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 121 LVATNIFGONTPALAATENHYAEMWADAAAFYGAATATATLPPFEAPEMTSAGG 179
QY 181 LLEQAAAVEASDTAAA-----NQLMNVVPAALQALQPTGCTTPSSKGLGLWKTVSP 233
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 180 TAAQSAVVVAAGAAAGSDDITAQLSLTPSLQSLA--TTATATSAG--WDTV-- 233
QY 234 HRSPISNMVMANNHMSMTNSGV---SMTNTLSSMLKGFAPAAAQAVQTAQNGVRAMS 290
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 234 -LQSIITILANLTGPGYSIILGLAIPGGMWLTFGQIL-GLAQNAPGVAAALGPKAAAGALS 291
QY 291 SLGSSLSG-----SGLGGVAAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTS 345
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 292 PLALRGYIADITPLGGGTGATARIYVGSLSVPGQWAEAPVVRVAVASVLPGTGAAP 351
QY 346 A-AERGGOMLGLPVQMGARAGGL-----SGVLRV 377
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
DQ 352 ALAAEAPGALFEGMALSSLAGALAGTAVRSAGAAARV 389

RESULT 10
OS3957 PRELIMINARY; PRT; 409 AA.
ID OS3957;
AC OS3957;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN Rv1808 OR MT1856.1 OR MTV049.30.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
```

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."; Nature 393:537-544 (1998).
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A. / Onkosh;
 RC STRAIN=CDC 1551 / Onkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."; Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AL022021; CAAL7729.1; -
 DR EMBL; AE007044; AAK46129.1; ALT_INIT.
 DR PIR; A70932; A70932.
 DR TIGR; MT1856.1; -
 DR TubercuList; Rv1808; -
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome.
 SQ SEQUENCE 409 AA; 39917 MW; 1E15202BACF36379 CRC64;
 Query Match 38.5%; Score 751; DB 16; Length 409;
 Best Local Similarity 44.7%; Pred. No. 8.4e-35;
 Matches 177; Conservative 53; Mismatches 138; Indels 28; Gaps 10;
 QY 2 VDFGALPEINSARMYAGPGSASLVAAQWDSVASFSAFQSVVWGLTVGWSIGS 61
 Db 1 MDFGALPEINSGRMYAGPGSGPLAARAAWDAALAEALYSAASYSGTIEGLTVPWMP 60
 QY 62 SAGLMVAASPYVAMSVTAGOAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMIL 121
 Db 61 SSITMAAAVAPYVAMISVTAGAEQAQAQAKIAAGVYETAFATVPPVIEANRALLMSL 120
 QY 122 IATNLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLPPFEAPBMTSAGGL 181
 Db 121 VATNIFGONTFAIAATEAHYAEWMAQDAAMYGAGSSATA-SQLAPFSEPPQTTFNSAT 179
 QY 182 LEQAAAVEASDTHAA-----NQLMNNVPQALQQLAQPTQGTTPSSKLGGLKTKTVSPH 234
 Db 180 AAQSAVVAQAAGAAASDDITAQLSQLISLLPSTLQSLA--TTATATSASAG--WDTV--- 232
 QY 235 RSPISNVSMNNHMTNSGV-----MTNTLSMLKGFAPAAQAQVTAQAQNGVRAM 289
 Db 233 LQSIITILANTGPYSIIIGLGAIPGGWHLTFQGLLAQNAQVGAALLGPKAAGALSPL 292
 QY 290 SSL-GSSLGS-SGLGGVAAANLGRAASVGLSVFPQAWAANAQVTPAARALPILTSLSA- 346
 Db 293 APLRGYIGDITPLGGGATGIAIRIYVGLSVFPQWAEAPVWRAVASVLPCTGAAPAL 352
 QY 347 AERPGQMLGCLPVQMGARAGGL-----SGVLRV 377
 Db 353 AAAPGALFGEMALSSLAGRALAGTAVRSGAARV 388
 RESULT 11
 Q7TZH7 PRELIMINARY; PRT; 409 AA.
 ID Q7TZH7
 AC Q7TZH7;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE32 OR MB1837.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Onkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."; Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AL022021; CAAL7729.1; -
 DR EMBL; AE007044; AAK46129.1; ALT_INIT.
 DR PIR; A70932; A70932.
 DR TIGR; MT1856.1; -
 DR TubercuList; Rv1808; -
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome.
 SQ SEQUENCE 409 AA; 39917 MW; 1E15202BACF36379 CRC64;
 Query Match 38.5%; Score 751; DB 16; Length 409;
 Best Local Similarity 44.7%; Pred. No. 8.4e-35;
 Matches 177; Conservative 53; Mismatches 138; Indels 28; Gaps 10;
 QY 2 VDFGALPEINSARMYAGPGSASLVAAQWDSVASFSAFQSVVWGLTVGWSIGS 61
 Db 1 MDFGALPEINSGRMYAGPGSGPLAARAAWDAALAEALYSAASYSGTIEGLTVPWMP 60
 QY 62 SAGLMVAASPYVAMSVTAGOAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMIL 121
 Db 61 SSITMAAAVAPYVAMISVTAGAEQAQAQAKIAAGVYETAFATVPPVIEANRALLMSL 120
 QY 122 IATNLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLPPFEAPBMTSAGGL 181
 Db 121 VATNIFGONTFAIAATEAHYAEWMAQDAAMYGAGSSATA-SQLAPFSEPPQTTFNSAT 179
 QY 182 LEQAAAVEASDTHAA-----NQLMNNVPQALQQLAQPTQGTTPSSKLGGLKTKTVSPH 234
 Db 180 AAQSAVVAQAAGAAASDDITAQLSQLISLLPSTLQSLA--TTATATSASAG--WDTV--- 232
 QY 235 RSPISNVSMNNHMTNSGV-----MTNTLSMLKGFAPAAQAQVTAQAQNGVRAM 289
 Db 233 LQSIITILANTGPYSIIIGLGAIPGGWHLTFQGLLAQNAQVGAALLGPKAAGALSPL 292
 QY 290 SSL-GSSLGS-SGLGGVAAANLGRAASVGLSVFPQAWAANAQVTPAARALPILTSLSA- 346
 Db 293 APLRGYIGDITPLGGGATGIAIRIYVGLSVFPQWAEAPVWRAVASVLPCTGAAPAL 352
 QY 347 AERPGQMLGCLPVQMGARAGGL-----SGVLRV 377
 Db 353 AAAPGALFGEMALSSLAGRALAGTAVRSGAARV 388
 RESULT 12
 Q7TZH8 PRELIMINARY; PRT; 399 AA.
 ID Q7TZH8
 AC Q7TZH8;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE31 OR MB1836.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis";
 RL EMBL; BX248340; CAD94540.1; -
 KW Complete proteome.
 SQ SEQUENCE 399 AA; 38840 MW; 1A0F437731BE74F2 CRC64;
 Query Match 38.0%; Score 740.5; DB 16; Length 399;

RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis";
 RL EMBL; BX248340; CAD94540.1; -
 KW Complete proteome.
 SQ SEQUENCE 409 AA; 39931 MW; D57892628B131A9E CRC64;
 Query Match 38.4%; Score 749; DB 16; Length 409;
 Best Local Similarity 44.3%; Pred. No. 1.1e-34;
 Matches 176; Conservative 51; Mismatches 140; Indels 30; Gaps 10;
 QY 2 VDFGALPEINSARMYAGPGSASLVAAQWDSVASFSAFQSVVWGLTVGWSIGS 61
 Db 1 MDFGALPEINSGRMYAGPGSGPLAARAAWDAALAEALYSAASYSGTIEGLTVPWMP 60
 QY 62 SAGLMVAASPYVAMSVTAGOAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMIL 121
 Db 61 SSITMAAAVAPYVAMISVTAGAEQAQAQAKIAAGVYETAFATVPPVIEANRALLMSL 120
 QY 122 IATNLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLPPFEAPBMTSAGGL 181
 Db 121 VATNIFGONTFAIAATEAHYAEWMAQDAAMYGAGSSATA-SQLAPFSEPPQTTFNSAT 179
 QY 182 LEQAAAVEASDTHAA-----NQLMNNVPQALQQLAQPTQGTTPSSKLGGLKTKTVSPH 234
 Db 180 AAQSAVVAQAAGAAASDDITAQLSQLISLLPSTLQSLA--TTATATSASAG--WDTV--- 232
 QY 235 RSPISNVSMNNHMTNSGV-----MTNTLSMLKGFAPAAQAQVTAQAQNGVRAM 291
 Db 233 LQSIITILANTGPYSIIIGLGAIPGGWHLTFQGLLAQNAQVGAALLGPKAAGALSPL 291
 QY 292 LGSLSGS-----SGLGGVAAANLGRAASVGLSVFPQAWAANAQVTPAARALPILTSLSA 346
 Db 292 LAPLGGYIADITPLGGGATGIAIRIYVGLSVFPQWAEAPVWRAVASVLPCTGAAPA 351
 QY 347 AERPGQMLGCLPVQMGARAGGL-----SGVLRV 377
 Db 352 LAAPGALFGEMALSSLAGRALAGTAVRSGAARV 388
 RESULT 12
 Q7TZH8 PRELIMINARY; PRT; 399 AA.
 ID Q7TZH8
 AC Q7TZH8;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE31 OR MB1836.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis";
 RL EMBL; BX248340; CAD94539.1; -
 KW Complete proteome.
 SQ SEQUENCE 399 AA; 38840 MW; 1A0F437731BE74F2 CRC64;
 Query Match 38.0%; Score 740.5; DB 16; Length 399;

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2004, 16:01:22 ; Search time 8830 Seconds
(without alignments)
11225.993 Million cell updates/sec

Title: US-09-597-796c-11

Perfect score: 2287
Sequence: 1 tctagaataatttggttta.....ggntgtacaaagccgaaa 2287

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Result No.	Score	Query Match	Length	DB ID	Description
1	2284	99.9	2287	6	BD251322 Fused pro
2	2284	99.9	2287	6	AR303127 Sequence
3	2284	99.9	2287	6	AR403735 Sequence
4	1770.8	77.4	1797	6	BD251333 Fused pro
5	1770.8	77.4	1797	6	AR403746 Sequence
6	1172.8	51.3	1801	6	BD251331 Fused pro
7	1172.8	51.3	1801	6	AR403744 Sequence
8	1169	51.1	1188	6	BD273807 Sequence
9	1169	51.1	1188	6	AX004989 Sequence
10	1169	51.1	1188	6	AR169205 Sequence
11	1169	51.1	1188	6	AR182495 Sequence
12	1169	51.1	1188	6	AR194878 Sequence
13	1169	51.1	1188	6	AR233150 Sequence
14	1169	51.1	1188	6	AR353355 Sequence
15	1169	51.1	1188	6	AX429698 Sequence
16	1169	51.1	1188	6	AX832683 Sequence
17	1169	51.1	1188	6	BD006378 Compounds
18	1169	51.1	1188	6	BD006498 Compounds
19	1169	51.1	1188	6	BD205870 Compounds
20	1169	51.1	1188	6	AE007000 Mycobacte
21	1169	51.1	1188	6	AX704275 Sequence
22	1169	51.1	1188	6	Bx842575 Mycobacte
23	1169	51.1	1188	6	BD273806 Sequence
24	1169	51.1	1188	6	AX004987 Sequence
25	1169	51.1	1188	6	BD069338 Compounds
26	1169	51.1	1188	6	EX248338 Mycobacte
27	1021	44.6	299450	1	EX248338 Mycobacte
28	955	41.8	20276	1	AE007161 Mycobacte
29	947.4	41.4	14029	1	AE007013 Mycobacte
30	942.6	41.2	348264	1	Bx842576 Mycobacte
31	941	41.1	3027	6	AR169207 Sequence
32	941	41.1	3027	6	AR182497 Sequence
33	941	41.1	3027	6	AR194880 Sequence
34	941	41.1	3027	6	AR233152 Sequence
35	941	41.1	3027	6	AR353357 Sequence
36	941	41.1	3027	6	AX429702 Sequence
37	941	41.1	3027	6	AX832687 Sequence
38	941	41.1	3027	6	BD06380 Compounds
39	941	41.1	3027	6	BD006500 Compounds
40	941	41.1	3027	6	BD205872 Compounds
41	923.8	40.4	316050	1	EX248346 Mycobacte
42	923.2	40.3	349606	15	Bx842583 Mycobacte
43	920.2	40.2	1179	6	AX750742 Sequence
44	884	38.7	1725	6	AR169206 Sequence
45	884	38.7	1725	6	AR182496 Sequence

ALIGNMENTS

RESULT 1	BD251322	BD251322	2287 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	Fused protein of Mycobacterium tuberculosis antigen and utilization thereof.					
DEFINITION	BD251322					
ACCESSION	BD251322	GI:33061092				
VERSION	JP 2002510494-A/1.					
KEYWORDS	synthetic construct					
SOURCE	synthetic construct					
ORGANISM	artificial sequences.					
REFERENCE	1 (bases 1 to 2287)					
AUTHORS	Skeiky, Y.A.W., Alderson, M. and Neto, A.C.					
TITLE	Fused protein of Mycobacterium tuberculosis antigen and utilization thereof					

Pred. No. is the number of results predicted by chance to have a

Db 601 TTCAAGTCGGTCTGGGGCTCTGACGGTGGGGTCTGGATAGGTTCGTCGGCGGTCTGA 660
QY 661 TGGTGGCGCGCCCTCGCGCTATGTGGGTGATGAGCGTCAACCGCGGGGAGCGCGAGC 720
Db 661 TGGTGGCGCGCCCTCGCGCTATGTGGGTGATGAGCGTCAACCGCGGGGAGCGCGAGC 720
QY 721 TGACCGCGCCCGCAGGTCTGGGGTGTGGGGGGGCTTACGAGCGCGGTATGGGTGAGCGG 780
Db 721 TGACCGCGCCCGCAGGTCTGGGGTGTGGGGGGGCTTACGAGCGCGGTATGGGTGAGCGG 780
QY 781 TGGCGCGCGCGGTATCGCGAGAACCGTGTCTGTAATCTGATAGCGACCAACC 840
Db 781 TGGCGCGCGCGGTATCGCGAGAACCGTGTCTGTAATCTGATAGCGACCAACC 840
QY 841 TCTTGGGGCAAAACCCCGCGGATCGCGGTCAACGAGCGCGAATACGCGAGATGTGGG 900
Db 841 TCTTGGGGCAAAACCCCGCGGATCGCGGTCAACGAGCGCGAATACGCGAGATGTGGG 900
QY 901 CCCAAGACCGCGCGGATTTGGCTACCGCGCGGACCGCGACGGCGACCGCGAGT 960
Db 901 CCCAAGACCGCGCGGATTTGGCTACCGCGCGGACCGCGACGGCGACCGCGAGT 960
QY 961 TGTCTCGGTTTCAGGAGGCGCGGAGATGACCGAGCGCGGTGGGTCTCTCGAGCAGGCGG 1020
Db 961 TGTCTCGGTTTCAGGAGGCGCGGAGATGACCGAGCGCGGTGGGTCTCTCGAGCAGGCGG 1020
QY 1021 CCGCGGTTCAGAGAGGCTCCGACACCGCGCGCGGTGGGTCTCTCGAGCAGGCGG 1080
Db 1021 CCGCGGTTCAGAGAGGCTCCGACACCGCGCGCGGTGGGTCTCTCGAGCAGGCGG 1080
QY 1081 AGCGGTGCAACAGCTGGCGCGCGGACCGCGGACCGCGCTTCTTCAAGCTGGGTG 1140
Db 1081 AGCGGTGCAACAGCTGGCGCGCGGACCGCGGACCGCGCTTCTTCAAGCTGGGTG 1140
QY 1141 GCCTGTGGAAGACGGTCTCCCGCATCGGTCCCGCATCAGCAACATGTGTGATGGGCA 1200
Db 1141 GCCTGTGGAAGACGGTCTCCCGCATCGGTCCCGCATCAGCAACATGTGTGATGGGCA 1200
QY 1201 ACAACACATGTGATGACCACTCGGTGTGTGATGACCAACACCTTGAGTCTGATGT 1260
Db 1201 ACAACACATGTGATGACCACTCGGTGTGTGATGACCAACACCTTGAGTCTGATGT 1260
QY 1261 TGAAGGGCTTTGTCTCGCGGGCGGCGCGGACCGGTGCAAAACCGCGGCGCAAAACGGGG 1320
Db 1261 TGAAGGGCTTTGTCTCGCGGGCGGCGCGGACCGGTGCAAAACCGCGGCGCAAAACGGGG 1320
QY 1321 TCGGGCGATGAGTCTCGTGGGAGCTCGCTGGGTCTTTCGGGTCTGGCGGTGGGGTGG 1380
Db 1321 TCGGGCGATGAGTCTCGTGGGAGCTCGCTGGGTCTTTCGGGTCTGGCGGTGGGGTGG 1380
QY 1381 CGCCCAACTTGGTTCGGCGGCTCGGTCTGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGG 1440
Db 1381 CGCCCAACTTGGTTCGGCGGCTCGGTCTGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGG 1440
QY 1441 CGGCCAACAGGAGTCAACCGCGCGGCGGCGGCGGTGCGGTGACCGCTGACCGAGCG 1500
Db 1441 CGGCCAACAGGAGTCAACCGCGCGGCGGCGGCGGTGCGGTGACCGCTGACCGAGCG 1500
QY 1501 CGCGGAAAGAGGCGCGGAGATCTCGCGGGGCTCGCGGTGGGGAGATGGGGCGCA 1560
Db 1501 CGCGGAAAGAGGCGCGGAGATCTCGCGGGGCTCGCGGTGGGGAGATGGGGCGCA 1560
QY 1561 GGGCGGTGTGGGCTCAGT 1620
Db 1561 GGGCGGTGTGTGGGCTCAGT 1620
QY 1621 ATTCTCGGAGCGCGGATATCGCGCGCGGCTTGTGCGAGGACCGGTTCGCGGACT 1680
Db 1621 ATTCTCGGAGCGCGGATATCGCGCGCGGCTTGTGCGAGGACCGGTTCGCGGACT 1680
QY 1681 TCCCGCGGTCTCGCGCGGTCTCGCATGTGTGCGCAAGTGGGGCGCACAGGTGTCA 1740

Db 1681 TCCCGCGGTCTCGCGCGGTCTCGCATGTGTGCGCAAGTGGGGCGCACAGGTGTCA 1740
QY 1741 ACATCAACACCAAACTGGGCTACAAACCGCGGTGGGCGCGGACCGGATCGTATCG 1800
Db 1741 ACATCAACACCAAACTGGGCTACAAACCGCGGTGGGCGCGGACCGGATCGTATCG 1800
QY 1801 ATCCCAACGGTGTCTGACCAAAACACCGGTGATCGGGCGGACCGGATCAATG 1860
Db 1801 ATCCCAACGGTGTCTGACCAAAACACCGGTGATCGGGCGGACCGGATCAATG 1860
QY 1861 CGTTCAAGCTCGGCTCGGCGCAAAACCTAGCGGTGATGTGTGCGGTATGACCGCACCC 1920
Db 1861 CGTTCAAGCTCGGCTCGGCGCAAAACCTAGCGGTGATGTGTGCGGTATGACCGCACCC 1920
QY 1921 AGGATGTGCGGTCTGACGCTCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 1980
Db 1921 AGGATGTGCGGTCTGACGCTCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 1980
QY 1981 GCGGCGTTCGGGTGAGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAG 2040
Db 1981 GCGGCGTTCGGGTGAGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAG 2040
QY 2041 GCGGCGTTCGGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAG 2100
Db 2041 GCGGCGTTCGGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAG 2100
QY 2101 CGCTGACCGGTTCGCGAAGACAGCATTTGAAACGGGTGATCCAGTTTCGATGCGCGCATCCAGC 2160
Db 2101 CGCTGACCGGTTCGCGAAGACAGCATTTGAAACGGGTGATCCAGTTTCGATGCGCGCATCCAGC 2160
QY 2161 CCGGTGATTCGGGCGGCGCGGTGCTGCTCAACGGCGGTAGGACAGGTGCTGCTGATGAAACAGG 2220
Db 2161 CCGGTGATTCGGGCGGCGCGGTGCTGCTCAACGGCGGTAGGACAGGTGCTGCTGATGAAACAGG 2220
QY 2221 CCGGTCTTAGGATATCCATCAGCTGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2280
Db 2221 CCGGTCTTAGGATATCCATCAGCTGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2280
QY 2281 CCGGAAA 2287
Db 2281 CCGGAAA 2287

RESULT 4
BD251333
LOCUS

DEFINITION Fused protein of Mycobacterium tuberculosis antigen and utilization thereof
ACCESSION BD251333
VERSION BD251333.1 GI:33061103
KEYWORDS JP 2002510494-A/12.
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 1797)
AUTHORS Skeiky, Y.A.W., Alderson, M. and Neto, A.C.
TITLE Fused protein of Mycobacterium tuberculosis antigen and utilization thereof
JOURNAL Patent: JP 2002510494-A 12 09-APR-2002;
CORIXA CORP
OS Artificial Sequence
PN JP 2002510494-A/12
PD 09-APR-2002
PF 07-APR-1999 JP 2000542460
PR 07-APR-1998 US 09/056556, 30-DEC-1998 US 09/223040 PI
C12N15/09, A61K39/04, A61K48/00, A61P31/04, C07K14/35, C07K19/00, PC
C12P21/02
CC C12N15/00
CC Description of Artificial Sequence: bi-fusion
protein TBH9-Ra35
CC (designated)
CC Mb59f)

BD251333 1797 bp DNA linear PAT 17-JUL-2003
Fused protein of Mycobacterium tuberculosis antigen and utilization thereof

FEATURES		FH		Key		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers		Location/Qualifiers	
----------	--	----	--	-----	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--	---------------------	--

RESULT 5
 AR403746
 LOCUS AR403746
 DEFINITION Sequence 25 from patent US 6627198.
 ACCESSION AR403746
 VERSION AR403746.1 GI:40151422
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED.
 REFERENCE 1 (bases 1 to 1797)
 AUTHORS Reed,S.G., Skeiky,Y.A., Dillon,D.C., Alderson,M. and Campos-Neto,A.

TITLE		Fusion proteins of Mycobacterium tuberculosis antigens and their uses	
JOURNAL		Patent: US 6627198-A 25 30-SEP-2003;	
FEATURES		Location/Qualifiers	
source		1..1797	
ORIGIN		/organism="unknown"	
		/mol_type="genomic DNA"	
Query Match		77.4%; Score 1770.8; DB 6; Length 1797;	
Best Local Similarity		99.9%; Pred. No. 1.7e-209;	
Matches 1772; Conservative		0; Mismatches 2; Indels 0; Gaps 0;	
QY	464	CATGTTGATTTCCGGGGGTTACCAACCGAGATCAACTCCGGAGAGTACGCGGGCC 523	
DB	24	CATGTTGATTTCCGGGGGTTACCAACCGAGATCAACTCCGGAGAGTACGCGGGCC 83	
QY	524	GGGTTCCGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTCACTTT 583	
DB	84	GGGTTCCGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTCACTTT 143	
QY	584	TTCCGGCGCGTTCAGTTCGGTTCAGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGG 643	
DB	144	TTCCGGCGCGTTCAGTTCGGTTCAGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGG 203	
QY	644	TTCCGTCCGGCGCTTCATGTTGGCGCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 703	
DB	204	TTCCGTCCGGCGCTTCATGTTGGCGCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCG 263	
QY	704	CGCGGGCAGCGCGCTGACCGCGCGCTGACCGCGCGCTGACCGCGCGCTGACCGCGCG 763	
DB	264	CGCGGGCAGCGCGCTGACCGCGCGCTGACCGCGCGCTGACCGCGCGCTGACCGCGCG 323	
QY	764	GGCGTATGGCTGACCGTCCCGCGCTGATCGCGAGAACCTGCTGTAATCATGAT 823	
DB	324	GGCGTATGGCTGACCGTCCCGCGCTGATCGCGAGAACCTGCTGTAATCATGAT 383	
QY	824	TCTGATAGCGACCACTCTTGGGCAAAACACCGCGCGATCGCGTCAACGAGCGCGA 883	
DB	384	TCTGATAGCGACCACTCTTGGGCAAAACACCGCGCGATCGCGTCAACGAGCGCGA 443	
QY	884	ATACGCGAGATGTGGCGCGCGTACAGCGCGCGTATGTTGCTAGCGCGCGCGCGCG 943	
DB	444	ATACGCGAGATGTGGCGCGCGTACAGCGCGCGTATGTTGCTAGCGCGCGCGCGCG 503	
QY	944	GACGGCGAGCGCGCTGCTGCGCTTCGAGGAGCGCGCGAGATGACGAGCGCGGTGG 1003	
DB	504	GACGGCGAGCGCGCTGCTGCGCTTCGAGGAGCGCGCGAGATGACGAGCGCGGTGG 563	
QY	1004	GCTCTCGAGCGCGCGCTGCGCTTCGAGGAGCGCGCGTTCGAGGAGCGCGGTGG 1063	
DB	564	GCTCTCGAGCGCGCGCTGCGCTTCGAGGAGCGCGCGTTCGAGGAGCGCGGTGG 623	
QY	1064	GATGACAAATGTCGCCCGCGCTGCAACAGCTGGCGCGCGCGCGCGCGCGCGCGCG 1123	
DB	624	GATGACAAATGTCGCCCGCGCTGCAACAGCTGGCGCGCGCGCGCGCGCGCGCGCG 683	
QY	1124	TTCTTCCAAAGCTGGTGGCTGGTGGAGAGCGTTCGCGCGCATCGCTTCGCGCATCAGCAA 1183	
DB	684	TTCTTCCAAAGCTGGTGGCTGGTGGAGAGCGTTCGCGCGCATCGCTTCGCGCATCAGCAA 743	
QY	1184	CATGTTGATGCGCAACCAACATGTCATGACCACTGGTGGTGGTGGTGGTGGTGGTGG 1243	
DB	744	CATGTTGATGCGCAACCAACATGTCATGACCACTGGTGGTGGTGGTGGTGGTGGTGG 803	
QY	1244	CACCTTGAGCTCGATGTTGAAGGGCTTGTCTCGCGCGCGCGCGCGCGCGCGCGCG 1303	
DB	804	CACCTTGAGCTCGATGTTGAAGGGCTTGTCTCGCGCGCGCGCGCGCGCGCGCGCG 863	
QY	1304	CGCGCGCAAAACGGGGTTCGGCGCATGAGCTGCTGGCGAGCTCGCTGGTTCCTCGGG 1363	
DB	864	CGCGCGCAAAACGGGGTTCGGCGCATGAGCTGCTGGCGAGCTCGCTGGTTCCTCGGG 923	

QY	1364	TCTGGCGGTGGGGTGGCGCGCAACTTGGGTCCGGCGGCTCGGTCCGTTCTGTTCTGGT 1423	
DB	924	TCTGGCGGTGGGGTGGCGCGCAACTTGGGTCCGGCGGCTCGGTCCGTTCTGTTCTGGT 983	
QY	1424	GCCTCAGGCTCGGGCGCGCGCAACAGAGTCAACCCGGCGCGCGCGCGCGCGCGTTCGGCT 1483	
DB	984	GCCTCAGGCTCGGGCGCGCGCAACAGAGTCAACCCGGCGCGCGCGCGCGCGTTCGGCT 1043	
QY	1484	GACAGCTGACAGCGCGCGCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGTTCGGCT 1543	
DB	1044	GACAGCTGACAGCGCGCGCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGTTCGGCT 1103	
QY	1544	GGGCGAGATGGGCGCGCAAGGCGCGGTTCAGTGGTTCAGTGGTTCAGTGGTTCAGTGGT 1603	
DB	1104	GGGCGAGATGGGCGCGCAAGGCGCGGTTCAGTGGTTCAGTGGTTCAGTGGTTCAGTGGT 1163	
QY	1604	ACCTATGATGCGCGCATTCCTCGGAGCGCGCGCATTCGCCCGCGCGCGCGCGCGTTCGCA 1663	
DB	1164	ACCTATGATGCGCGCATTCCTCGGAGCGCGCGCATTCGCCCGCGCGCGCGCGTTCGCA 1223	
QY	1664	GGACCGGTTCGGCGACTTCCCGCGCTGCGCGCTCGACCGTTCGCGCATGCTCGCCCAAGT 1723	
DB	1224	GGACCGGTTCGGCGACTTCCCGCGCTGCGCGCTCGACCGTTCGCGCATGCTCGCCCAAGT 1283	
QY	1724	GGGCGCAAGTGGTTCAGATCAACACCAACTGGGCTACAAACAGCGCGTGGCGCGCG 1783	
DB	1284	GGGCGCAAGTGGTTCAGATCAACACCAACTGGGCTACAAACAGCGCGTGGCGCGCG 1343	
QY	1784	GACCGCATCGTCAATCGATCCCAACCGTTCGCTGACCAACCAACCAACCGTTCGCGG 1843	
DB	1344	GACCGCATCGTCAATCGATCCCAACCGTTCGCTGACCAACCAACCAACCGTTCGCGG 1403	
QY	1844	GGCACCGCATCGTTCAGGTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1903	
DB	1404	GGCACCGCATCGTTCAGGTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1463	
QY	1904	CGGTTGATGCGCGACCGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1963	
DB	1464	CGGTTGATGCGCGACCGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1523	
QY	1964	GTCGCGCGCATCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2023	
DB	1524	GTCGCGCGCATCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1583	
QY	2024	CGTGGCGAGCGCGCGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2083	
DB	1584	CGTGGCGAGCGCGCGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1643	
QY	2084	CGTGGCGAGCGCGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2143	
DB	1644	CGTGGCGAGCGCGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1703	
QY	2144	CGATGCGCGCATTCAGCGCGGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2203	
DB	1704	CGATGCGCGCATTCAGCGCGGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1763	
QY	2204	GGTGGTATGACACCGCGCGTTCCTAGGATATC 2237	
DB	1764	GGTGGTATGACACCGCGCGTTCCTAGGATATC 1797	

RESULT 6	
BD251331	
LOCUS	1801 bp DNA linear PAT 17-JUL-2003
DEFINITION	Fused protein of Mycobacterium tuberculosis antigen and utilization thereof.
ACCESSION	BD251331
VERSION	BD251331.1 GI:33061101
KEYWORDS	JP 2002510494-A/10.
SOURCE	synthetic construct
ORGANISM	synthetic construct
REFERENCE	1 (bases 1 to 1801)

Db 371 TTCTGATAGCACCACCTCTTGGGGGAAACACACCCCGCGGATCGCGGTCAACGAGGCGG 430
QY 893 AATACGGCAGATGTGGGCCCAAGACGCGCGCGATGTTTGGCTACGCGCGCGGACGCG 942
Db 431 AATACGGCAGATGTGGGCCCAAGACGCGCGCGATGTTTGGCTACGCGCGCGGACGCG 490
QY 943 CGACGGCAGCGGACGCTGTGCTGCGTTCGAGGAGGCGCGGAGATGACAGCGCGGCTG 1002
Db 491 CGACGGCAGCGGACGCTGTGCTGCGTTCGAGGAGGCGCGGAGATGACAGCGCGGCTG 550
QY 1003 GGCTCTCTGACGAGCGCGCGCTGCGAGGAGGCTTCCGACACCGCGCGCGGAAACCACT 1062
Db 551 GGCTCTCTGACGAGCGCGCGCTGCGAGGAGGCTTCCGACACCGCGCGCGGAAACCACT 610
QY 1063 TGATGAACATGTGCGCGCGCGCTGCAACAGCTGGCGCCAGCCACGAGGCGACACGCG 1122
Db 611 TGATGAACATGTGCGCGCGCGCTGCAACAGCTGGCGCCAGCCACGAGGCGACACGCG 670
QY 1123 CTTCTTCCAACTGGTGGCGCTGTGGAAGAGGCTCTCGCGCATCGGTTCGCGCATCAGCA 1182
Db 671 CTTCTTCCAACTGGTGGCGCTGTGGAAGAGGCTCTCGCGCATCGGTTCGCGCATCAGCA 730
QY 1183 ACATGGTGTGATGCCAACACACATGTCGATGACCAACTCGGGTGTGTCGATGACCA 1242
Db 731 ACATGGTGTGATGCCAACACACATGTCGATGACCAACTCGGGTGTGTCGATGACCA 790
QY 1243 ACACCTTGAGCTCGATGTGAGGGCTTGTCTCGGGGCGGCGCGCGCGCTGCAGAA 1302
Db 791 ACACCTTGAGCTCGATGTGAGGGCTTGTCTCGGGGCGGCGCGCGCGCTGCAGAA 850
QY 1303 CCGCGGCGGCGGCGGCTCGGCGGATGAGCTCGCTGGCGAGCTCGCTGGGTTCCTCGG 1362
Db 851 CCGCGGCGGCGGCGGCTCGGCGGATGAGCTCGCTGGCGAGCTCGCTGGGTTCCTCGG 910
QY 1363 GTCTGGCGGCTGGGCTGGCGGCGGCTGAGCTGGGTGGCGGCGGCTCGGTGCTGCTGG 1422
Db 911 GTCTGGCGGCTGGGCTGGCGGCGGCTGAGCTGGGTGGCGGCGGCTCGGTGCTGCTGG 970
QY 1423 TGCGCGGAGGCTGGGCGGCGGCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1482
Db 971 TGCGCGGAGGCTGGGCGGCGGCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1030
QY 1483 TGACGAGCTGACGAGCGCGGCGGCGGCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1542
Db 1031 TGACGAGCTGACGAGCGCGGCGGCGGCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1090
QY 1543 TGGGCGAGATGGGCGGCGGCGGCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1602
Db 1091 TGGGCGAGATGGGCGGCGGCGGCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1150
QY 1603 GACCTATGTATGCGCGCATTTCTCGGCGGCGGCGGCGGCA 1639
Db 1151 GACCTATGTATGCGCGCATTTCTCGGCGGCGGCGGCGGCA 1187

RESULT 9
AX004989
LOCUS AX004989 1188 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 108 from Patent WO9909186.
ACCESSION AX004989
VERSION AX004989.1 GI:9928350

KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium;
tuberculosis complex.

REFERENCE
1 Portnoi, D. and Guigueno, A.
AUTHORS Polypeptide nucleic sequences exported from mycobacteria, vectors
TITLE comprising same and uses for diagnosing and preventing tuberculosis
JOURNAL Patent: WO 9909186-A 108 25-FEB-1999;
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)

FEATURES
source
CDS
Location/Qualifiers
1..1188
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
->_1188
/note="unnamed protein product; SEQ ID NO 5P"
/codon_start=1
/transl_table=11
/protein_id="CAC04959.1"
/db_xref="GI:9928351"
/translation="GHVYVDFGALPPRINSARMYAGPSASLVAALAAQWDSVASDLPS
AASAFQSVVWGLTVGSGAGLMVAASPYVAWNSVTAGQAELETAQVRVAAAYE
TAGVLTVPVIAENRAELMILIAIATLQNTPAIVNEABYGEWMAQDAAMEGYAA
ATATATALLPFEAPEATSALEAAVEASDTAAANQLMNYPQALQQAQPT
QGTTPSSKLGLWKTVPFHRSPISNMVSMANNHMTNSGVSMNTLSMLKFAFAA
AAAVOTAAQNGVRASISLSSGLGGVAAALGRAASVGLSVPOAWAANQAV
TPAARALPLTSLTSAERGPQMLGGLPVGQMGARAGGLSGVLVFPFPRVMPHSPA
AG"

ORIGIN
Query Match 51.1%; Score 1169; DB 6; Length 1188;
Best Local Similarity 99.6%; Pred. No. 3.4e-135;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 463 TCATGGTGGATTCGGGGCGTTACACCGGAGATCAACTCCGCGAGGATGTACCCCGGCC 522
Db 11 TAATGGTGGATTCGGGGCGTTACACCGGAGATCAACTCCGCGAGGATGTACCCCGGCC 70
QY 523 CGGGTTCCGCGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGT 582
Db 71 CGGGTTCCGCGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGT 130
QY 583 TTTCCGCGCGCTCGCGGCTTTCAGTGGTGGTCTGGGGTCTGACCGTGGGGTCTGGGATG 642
Db 131 TTTCCGCGCGCTCGCGGCTTTCAGTGGTGGTCTGGGGTCTGACCGTGGGGTCTGGGATG 190
QY 643 GTTCGCGCGCGCTCGCTGGTGGCGCGCTCGCGTATGCGGTGATGAGCTGCA 702
Db 191 GTTCGCGCGCGCTCGCTGGTGGCGCGCTCGCGTATGCGGTGATGAGCTGCA 250
QY 703 CCGCGGCGCAGCGCGAGCTGACCGCGCGCGCGCTGCGGGTTGCTGCGCGCGCTTACGAGA 762
Db 251 CCGCGGCGCAGCGCGAGCTGACCGCGCGCGCGCTGCGGGTTGCTGCGCGCGCTTACGAGA 310
QY 763 CCGCGTATGGGCTGACCGTGGCGCGCGCGCTGATGCGCGAGAACCGTGAACCTGATGA 822
Db 311 CCGCGTATGGGCTGACCGTGGCGCGCGCGCTGATGCGCGAGAACCGTGAACCTGATGA 370
QY 823 TTCTGATAGCACCACCACTCTTGGGGCAAAACACACCCCGCGGATCGCGGTCAACGAGGCG 882
Db 371 TTCTGATAGCACCACCACTCTTGGGGCAAAACACACCCCGCGGATCGCGGTCAACGAGGCG 430
QY 883 AATACGGCAGATGTGGGCCCAAGACGCGCGCGGATGTTGGCTACGCGCGCGGACGCG 942
Db 431 AATACGGCAGATGTGGGCCCAAGACGCGCGCGGATGTTGGCTACGCGCGCGGACGCG 490
QY 943 CGACGGCAGCGGCGGCTGCTGCGGCTTCGAGGAGGCGCGGAGATGACAGCGCGGCTG 1002
Db 491 CGACGGCAGCGGCGGCTGCTGCGGCTTCGAGGAGGCGCGGAGATGACAGCGCGGCTG 550
QY 1003 GGCTCTCTGACGAGCGCGCGCTGAGGAGGCTTCCGACACCGCGCGCGGAAACCACT 1062
Db 551 GGCTCTCTGACGAGCGCGCGCTGAGGAGGCTTCCGACACCGCGCGCGGAAACCACT 610
QY 1063 TGATGAACATGTGCGCGCGCGCTGCAACAGCTGGCGCCAGCCACGAGGCGACACGCG 1122
Db 611 TGATGAACATGTGCGCGCGCGCTGCAACAGCTGGCGCCAGCCACGAGGCGACACGCG 670
QY 1123 CTTCTTCCAACTGGTGGCGCTGTGGAAGAGGCTCTCGCGCATCGGTTCGCGCATCAGCA 1182
Db 671 CTTCTTCCAACTGGTGGCGCTGTGGAAGAGGCTCTCGCGCATCGGTTCGCGCATCAGCA 730

QY	1183	ACATGGTGTTCGATGGCGCAACCAACCATGTCGATGACCAACTCGGGTGTGTCGATGACCA	1242
Db	731	ACATGGTGTTCGATGGCGCAACCAACCATGTCGATGACCAACTCGGGTGTGTCGATGACCA	790
QY	1243	ACACCTTGAAGTTCGATGTTGAAGGGCTTTGCTCCCGCGCGCGCCGCCAGGCGCGTGCAAA	1302
Db	791	ACACCTTGAAGTTCGATGTTGAAGGGCTTTGCTCCCGCGCGCGCCGCCAGGCGCGTGCAAA	850
QY	1303	CGCGCGCGCAAAACGGGGTTCGGGCGATGAGTCTCGTGGGCGAGCTCGCTGGGTTCTTCGG	1362
Db	851	CGCGCGCGCAAAACGGGGTTCGGGCGATGAGTCTCGTGGGCGAGCTCGCTGGGTTCTTCGG	910
QY	1363	GTCTGGGCGGTGGGTGGCGGCCCACTTGGGTGGCGCGCGCTTCGTTTCGTTGTCGG	1422
Db	911	GTCTGGGCGGTGGGTGGCGGCCCACTTGGGTGGCGCGCGCTTCGTTTCGTTGTCGG	970
QY	1423	TGCGGAGGCTTGGGCGCGCGCAACCAAGGAGTCAACCGCGCGCGCGCGCGCGCGCGCG	1482
Db	971	TGCGGAGGCTTGGGCGCGCGCAACCAAGGAGTCAACCGCGCGCGCGCGCGCGCGCGCG	1030
QY	1483	TGACGAGCTTACCAGCGCGCGGGAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1542
Db	1031	TGACGAGCTTACCAGCGCGCGGGAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1090
QY	1543	TGCGGCGAGTGGCG	1602
Db	1091	TGCGGCGAGTGGCG	1150
QY	1603	GACCTTATGTGATGCGCGATTCCTCGGCGAGCGCGGA	1639
Db	1151	GACCTTATGTGATGCGCGATTCCTCGGCGCGCGGCTA	1187

RESULT 10

AR169205

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches 1172; Conservative

51.1%; Score 1169; DB 6; Length 3058;

99.6%; Pred. No. 2.7e-135;

0; Mismatches 5; Indels 0; Gaps 0;

QY

463

TCATGGTGGATTTCCGGGCGGTACACCGAGATCACTCCGCGAGGATGACCGCGCC

522

Db

391

TAATGGTGGATTTCCGGGCGGTACACCGAGATCACTCCGCGAGGATGACCGCGCC

450

QY

523

CGGGTTCCGGCTCGCTGGTGGCGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGT

582

Db

451

CGGGTTCCGGCTCGCTGGTGGCGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGT

510

QY

583

TTTCGGCGCGGTGGGTTTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG

642

Db

511

TTTCGGCGCGGTGGGTTTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG

570

QY

643

GTTCGTGGCGGGTCTGATGGTGGCGCGCGCTCGCGCTGATGTGGCGTGGATGAGCGTCA

702

RESULT 11
AR182495
LOCUS
DEFINITION
ACCESSION

AR182495
Sequence 101 from patent US 6338852.
AR182495

linear
3058 bp
DNA
PAT 20-APR-2002

Db	571	GTTCGTGGCGGGTCTGATGTTGGCGGCGCTCGCGCTATGTTGGCTGATGAGCGTCA	630
QY	703	CCGCGGGGCAAGGCGAGCTGACCGCGCCCGAGTCCGGTTCGGTCCGGCGCGCTACGAGA	762
Db	631	CCGCGGGGCAAGGCGAGCTGACCGCGCCCGAGTCCGGTTCGGTCCGGCGCGCTACGAGA	690
QY	763	CGCGTATGGCTGACCGGTGCTCCCGCGCGGTGATCGCGGAGAACCGTGTGAACTGATGA	822
Db	691	CGCGTATGGCTGACCGGTGCTCCCGCGCGGTGATCGCGGAGAACCGTGTGAACTGATGA	750
QY	823	TTCTGATAGCGACCAACTTTTGGGGCAAAACACCCCGCGGATCGCGGTCAAAGAGCCG	882
Db	751	TTCTGATAGCGACCAACTTTTGGGGCAAAACACCCCGCGGATCGCGGTCAAAGAGCCG	810
QY	883	AATACGCGAGATGTGGGCGCCCAAGAGCGCGCGCGATGTTTGGCTACGCGCGCGCGCG	942
Db	811	AATACGCGAGATGTGGGCGCCCAAGAGCGCGCGCGATGTTTGGCTACGCGCGCGCGCG	870
QY	943	CGAGCGCGAGCGGCGACCTTCTGCGTTCCAGAGAGCGCGCGGAGATGACACAGCGCGGT	1002
Db	871	CGAGCGCGAGCGGCGACCTTCTGCGTTCCAGAGAGCGCGCGGAGATGACACAGCGCGGT	930
QY	1003	GGCTCTCTGAGCAGCGCGCGCGGTTCGAGAGGCTTCGACACCGCGCGCGCGGCAACCGT	1062
Db	931	GGCTCTCTGAGCAGCGCGCGCGGTTCGAGAGGCTTCGACACCGCGCGCGCGGCAACCGT	990
QY	1063	TGATGAACAATGTGCCCCAGCGCGTGCACACAGTGTGGCCAGCCACACAGCGCGCGCG	1122
Db	991	TGATGAACAATGTGCCCCAGCGCGTGCACACAGTGTGGCCAGCCACACAGCGCGCGCG	1050
QY	1123	CTTCTTCCAAAGCTGGGTGGCGTGTGGAAGACGGTCTCGCGCGATCGGTTCGCGGATGACCA	1182
Db	1051	CTTCTTCCAAAGCTGGGTGGCGTGTGGAAGACGGTCTCGCGCGATCGGTTCGCGGATGACCA	1110
QY	1183	ACATGGTGTTCGATGCGCAACCAACATGTCGATGACCAACTCGGGTGTGTCGATGACCA	1242
Db	1111	ACATGGTGTTCGATGCGCAACCAACATGTCGATGACCAACTCGGGTGTGTCGATGACCA	1170
QY	1243	ACACCTTGAAGTTCGATGTTGAAGGGCTTTGCTCCGCGCGCGCGCGCGCGCGCGCG	1302
Db	1171	ACACCTTGAAGTTCGATGTTGAAGGGCTTTGCTCCGCGCGCGCGCGCGCGCGCGCG	1230
QY	1303	CCGCGCGCAAAACGGGGTCCGGCGGATGAGTCTCGTGGCGAGCTCGCTGGGTTCTTCGG	1362
Db	1231	CCGCGCGCAAAACGGGGTCCGGCGGATGAGTCTCGTGGCGAGCTCGCTGGGTTCTTCGG	1290
QY	1363	GTCTGGGCGGTGGGTGGCGCGCAACTTGGGTTCGGGCGCGCTCGGTTCGTTGTCGG	1422
Db	1291	GTCTGGGCGGTGGGTGGCGCGCAACTTGGGTTCGGGCGCGCTCGGTTCGTTGTCGG	1350
QY	1423	TGCGCGAGGCTGGGCGCGCGCGCAACAGGAGTCAACCGCGCGCGCGCGCGCGCGCG	1482
Db	1351	TGCGCGAGGCTGGGCGCGCGCGCAACAGGAGTCAACCGCGCGCGCGCGCGCGCGCG	1410
QY	1483	TGACCGGCTGACAGCGCGCGGGAAGAGGCGCGCGGAGATGCTGGCGCGGTTCGCG	1542
Db	1411	TGACCGGCTGACAGCGCGCGGGAAGAGGCGCGCGGAGATGCTGGCGCGGTTCGCG	1470
QY	1543	TGGGCGAGATGGGCGCGCGCGCGCGCGTGGGTTCAGTGGTGTGCTGCGTTCGCGCG	1602
Db	1471	TGGGCGAGATGGGCGCGCGCGCGCGCGTGGGTTCAGTGGTGTGCTGCGTTCGCGCG	1530
QY	1603	GACCTTATGATGCGCGATTCCTCGGCGCGCGCGGA	1639
Db	1531	GACCTTATGATGCGCGATTCCTCGGCGCGCGCGCTA	1567

VERSION	AR182495.1	GI:20225702
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	1 (bases 1 to 3058)	
AUTHORS	Reed,S.G., Skeiky,Y.A.W., Dillon,D.C., Campos-Neto,A., Houghton,R., Vedwick,T.S. and Twardzik,D.R.	
TITLE	Compounds and methods for diagnosis of tuberculosis	
JOURNAL	Patent: US 6338852-A 101 15-JAN-2002;	
FEATURES	Location/Qualifiers	
source	1..3058	
	/organism="unknown"	
ORIGIN	/mol_type="unassigned DNA"	
Query Match	51.1%; Score 1169; DB 6; Length 3058;	
Best Local Similarity	99.6%; Pred. No. 2.7e-135;	
Matches 1172; Conservative	0; Mismatches 5; Indels 0; Gaps 0;	
QY	463 TCATGGTGAATTCGGGCGGTTACACCGAGATCAACTCCGCGAGATGATAGCGCGCC	522
DB	391 TAATGGTGAATTCGGGCGGTTACACCGAGATCAACTCCGCGAGATGATAGCGCGCC	450
QY	523 CGGGTTCCGCGCTCGCTGGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGA	582
DB	451 CGGGTTCCGCGCTCGCTGGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGA	510
QY	1003 GGCTCTCGAGCGCGCGCGCTCGAGGAGCGCTCCGACACCGCGCGGGAACCACTG	1062
DB	931 GGCTCTCGAGCGCGCGCGCTCGAGGAGCGCTCCGACACCGCGCGGGAACCACTG	990
QY	1063 TGATGAACAATGCCCCAGCGCTGCAACAGCTGGCGCCAGCCACGAGGACCAACGC	1122
DB	991 TGATGAACAATGCCCCAGCGCTGCAACAGCTGGCGCCAGCCACGAGGACCAACGC	1050
QY	1123 CTTCTTCAAGCTGGGTGGCGCTGGAAGAGCGGTCTCGCGCATCGGTTCGCGCAT	1182
DB	1051 CTTCTTCAAGCTGGGTGGCGCTGGAAGAGCGGTCTCGCGCATCGGTTCGCGCAT	1110
QY	1183 ACATGGTTCATGGCCCAACACCAATGTCATGACCAACTCGGTGTGTCATGACCA	1242
DB	1111 ACATGGTTCATGGCCCAACACCAATGTCATGACCAACTCGGTGTGTCATGACCA	1170
QY	1243 ACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCGCGCGCGCGCCGACGAGCGCTG	1302
DB	1171 ACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCGCGCGCGCGCCGACGAGCGCTG	1362
QY	1303 CGCGCGCGCAAAACCGGGTCCGGGCGATGAGCTCGCTGGGACAGCTCGCTGGGTTCTTC	1290
DB	1231 CGCGCGCGCAAAACCGGGTCCGGGCGATGAGCTCGCTGGGACAGCTCGCTGGGTTCTTC	1290
QY	1363 GTCTGGGCGGTGGGGTGGCGCGCAACTTGGGTGGGGCGGCTCGTGGGTTCGTTCTCG	1422
DB	1291 GTCTGGGCGGTGGGGTGGCGCGCAACTTGGGTGGGGCGGCTCGTGGGTTCGTTCTCG	1350
QY	1423 TGCGCAGCGCTGGGCGCGCGCAACAGCAGTCAACCGCGCGCGCGCGCGCTCGCGC	1482
DB	1351 TGCGCAGCGCTGGGCGCGCGCAACAGCAGTCAACCGCGCGCGCGCGCGCTCGCGC	1410
QY	1483 TGACACGCTGACACAGCGCGCGCAAGAGAGGCGCGGCGAGATGCTGGGCGGCTCGCGC	1542
DB	1411 TGACACGCTGACACAGCGCGCGCAAGAGAGGCGCGGCGAGATGCTGGGCGGCTCGCGC	1470
QY	1543 TGGGGCAGATGGGCGCGCAGGCGGCTGGTGGGCTCAGTGGTGGTGGTGGTGGTGGT	1602
DB	1471 TGGGGCAGATGGGCGCGCAGGCGGCTGGTGGGCTCAGTGGTGGTGGTGGTGGTGGT	1530
QY	1603 GACCTATGTGATCCGCAATTCCTCCGCGAGCGCGCA 1639	
DB	1531 GACCTATGTGATCCGCAATTCCTCCGCGAGCGCGCTA 1567	
RESULT 12		
AR194878		
LOCUS	AR194878 3058 bp DNA linear PAT 20-APR-2002	
DEFINITION	Sequence 106 from patent US 6350456.	
ACCESSION	AR194878	
VERSION	AR194878.1 GI:20244315	
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unclassified.	
REFERENCE	1 (bases 1 to 3058)	
AUTHORS	Reed,S.G., Skeiky,Y.A.W. and Dillon,D.C.	
TITLE	Compositions and methods for the prevention and treatment of M. tuberculosis infection	
JOURNAL	Patent: US 6350456-A 106 26-FEB-2002;	
FEATURES	Location/Qualifiers	
source	1..3058	
	/organism="unknown"	
ORIGIN	/mol_type="unassigned DNA"	
Query Match	51.1%; Score 1169; DB 6; Length 3058;	
Best Local Similarity	99.6%; Pred. No. 2.7e-135;	
Matches 1172; Conservative	0; Mismatches 5; Indels 0; Gaps 0;	
QY	463 TCATGGTGAATTCGGGCGGTTACACCGAGATCAACTCCGCGAGATGATAGCGCGCC	522
DB	391 TAATGGTGAATTCGGGCGGTTACACCGAGATCAACTCCGCGAGATGATAGCGCGCC	450
QY	523 CGGGTTCCGCGCTCGCTGGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGA	582
DB	451 CGGGTTCCGCGCTCGCTGGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGA	510
QY	583 TTTCCGCGCGCGCTCGCGTTTCAGTCGGTGGTCTGGGCTGACAGGTGGGTGGTGGATAG	642
DB	511 TTTCCGCGCGCGCTCGCGTTTCAGTCGGTGGTCTGGGCTGACAGGTGGGTGGTGGATAG	570
QY	643 GTTCTGTCGCGGCTCGATGTGGCGCGGCTCGCGTATGTCGCGTATGTCGATGACG	702
DB	571 GTTCTGTCGCGGCTCGATGTGGCGCGGCTCGCGTATGTCGCGTATGTCGATGACG	630
QY	703 CGCGGGGACGCGAGCTGACCGCGCGCCAGGTTCGGGTGGTGGCGCGCTTACGAGA	762
DB	631 CGCGGGGACGCGAGCTGACCGCGCGCCAGGTTCGGGTGGTGGCGCGCTTACGAGA	690
QY	763 CGCGTATGGCTGACGCTGCGCGCGCGGTGATGCGCGAGACCGTGGTGAATGATCA	822
DB	691 CGCGTATGGCTGACGCTGCGCGCGCGGTGATGCGCGAGACCGTGGTGAATGATCA	750
QY	823 TTCTGATAGCAACCACTCTTGGGGGCAAAACACCCCGCGGATCGCGGTCAACGAGCGG	882
DB	751 TTCTGATAGCAACCACTCTTGGGGGCAAAACACCCCGCGGATCGCGGTCAACGAGCGG	810
QY	883 AATACGGGAGATGTGGGCGCAAGCGCGCGCGATGTTGGTGTACCGCGCGCGGACGG	942
DB	811 AATACGGGAGATGTGGGCGCAAGCGCGCGCGATGTTGGTGTACCGCGCGCGGACGG	870
QY	943 CGACGCGCACGCGACGTTGCTGCGCTTCGAGGAGGCGCGGAGATGACACGCGCGGTG	1002
DB	871 CGACGCGCACGCGACGTTGCTGCGCTTCGAGGAGGCGCGGAGATGACACGCGCGGTG	930
QY	1003 GGCTCTCGAGCGCGCGCGCTCGAGGAGCGCTCCGACACCGCGCGGGAACCACTG	1062
DB	931 GGCTCTCGAGCGCGCGCGCTCGAGGAGCGCTCCGACACCGCGCGGGAACCACTG	990
QY	1063 TGATGAACAATGCCCCAGCGCTGCAACAGCTGGCGCCAGCCACGAGGACCAACGC	1122
DB	991 TGATGAACAATGCCCCAGCGCTGCAACAGCTGGCGCCAGCCACGAGGACCAACGC	1050
QY	1123 CTTCTTCAAGCTGG	

763 CGCGCTATGGCTGAGGTGCCCCCGCGGTGATCGCCGAGAACCGTCTGAATGATGA 822
Db CGCGCTATGGCTGAGGTGCCCCCGCGGTGATCGCCGAGAACCGTCTGAATGATGA 750
QY TTCTGATGACGACCAACCTCTTTGGGGCAAAAACACCCCGCGATCGCGGTCAACGAGGCGG 882
Db TTCTGATGACGACCAACCTCTTTGGGGCAAAAACACCCCGCGATCGCGGTCAACGAGGCGG 810
QY AATACGGCAGATGTGGGCCCCAAGACCGCCCGCGCGATGTTGGCTACGCCGCGGCGAGG 942
Db AATACGGCAGATGTGGGCCCCAAGACCGCCCGCGCGATGTTGGCTACGCCGCGGCGAGG 870
QY CGACGGCAGCGGAGCGGTGCTGCTGCGGTTCGAGAGGCGCGGAGATGACACGCGCGGCTG 1002
Db CGACGGCAGCGGAGCGGTGCTGCTGCGGTTCGAGAGGCGCGGAGATGACACGCGCGGCTG 930
QY GGCTCTCAGACGAGCGCGCGGTGCGAGGAGGCTTCCACACCGCGCGGCGGCAACAGT 1062
Db GGCTCTCAGACGAGCGCGCGGTGCGAGGAGGCTTCCACACCGCGCGGCGGCAACAGT 990
QY TGATGAACATGTGCCCGAGGCGCTGCAACAGCTGGCCAGCCACGACGAGGACACACG 1122
Db TGATGAACATGTGCCCGAGGCGCTGCAACAGCTGGCCAGCCACGACGAGGACACACG 1050
QY CTCTTTCCAGCTGGGTGCGCTGTGGAAGACCGTCTCGCCGATCGTTCGCCGATCAGCA 1182
Db CTCTTTCCAGCTGGGTGCGCTGTGGAAGACCGTCTCGCCGATCGTTCGCCGATCAGCA 1110
QY ACATGCTGATGCGCAACACCAACATGTCGATGACCAACTCGGTGTCGATGACCA 1242
Db ACATGCTGATGCGCAACACCAACATGTCGATGACCAACTCGGTGTCGATGACCA 1170
QY ACACCTTTGAGCTCGATGTTGAGGGCTTTGCTCGGGCGCGCCCGCAGGCGCTGCAAA 1302
Db ACACCTTTGAGCTCGATGTTGAGGGCTTTGCTCGGGCGCGCCCGCAGGCGCTGCAAA 1230
QY CGCGCGCAAAACGCGGTTCGCGCGATGAGTCTGCTGCGCGGCGCGCGCGCGCTGCGG 1362
Db CGCGCGCAAAACGCGGTTCGCGCGATGAGTCTGCTGCGCGGCGCGCGCGCGCTGCGG 1290
QY GTCTGGCGGTGGGTGGCGCGCACTTGGGTGGCGCGCGCGCGCTGGTCTGGTCTGG 1422
Db GTCTGGCGGTGGGTGGCGCGCACTTGGGTGGCGCGCGCGCGCTGGTCTGGTCTGG 1350
QY TGCGCGAGCGCTGGCGCGCGCAACACGAGCAGTCAACCGCGCGCGCGCGCGCTGCGCG 1482
Db TGCGCGAGCGCTGGCGCGCGCAACACGAGCAGTCAACCGCGCGCGCGCGCGCTGCGCG 1410
QY TGACAGCTGACAGCGCGCGGAAAGAGGCGCGCGCGAGATGCTGGCGCGCGCTGCGCG 1542
Db TGACAGCTGACAGCGCGCGGAAAGAGGCGCGCGCGAGATGCTGGCGCGCGCTGCGCG 1470
QY TGGGCGAGATGGCGCGCGGTGGTGGGTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1602
Db TGGGCGAGATGGCGCGCGGTGGTGGGTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1530
QY GACCTATGATGCGCGATTCCTCGGCGAGCGGCGA 1639
Db GACCTATGATGCGCGATTCCTCGGCGAGCGGCGTA 1567

RESULT 13

AR233150 LOCUS 3058 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 101 from patent US 6458366.
ACCESSION AR233150
VERSION AR233150.1 GI:27275586
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3058)
AUTHORS Reed,S.G., Skeiky,Y.A.W., Dillon,D.C., Campos-Neto,A., Houghton,R.,

Vedvick,T.S., Twardzik,D.R., Lodes,M.J. and Hendrickson,R.C.
Compounds and methods for diagnosis of tuberculosis
Patent: US 6458366-A 101 01-OCI-2002;
Location/Qualifiers
1. 3058
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 51.1%; Score 1169; DB 6; Length 3058;
Best Local Similarity 99.6%; Pred. No. 2.7e-135;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 463 TCATGTTGGATTTTCGGGGCGTTACACCGGAGATCAACTCCGGAGGATGTAGCCCGGCC 522
Db 391 TAATGTTGGATTTTCGGGGCGTTACACCGGAGATCAACTCCGGAGGATGTAGCCCGGCC 450
QY 523 GGGTTTCGGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGGTGGCGAGTACCTGT 582
Db 451 CGGTTTCGGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGGTGGCGAGTACCTGT 510
QY 583 TTTCCGGCGCGTTCAGTTCGGTGTCTGGGCTCTGAGGTGGGTCTGGGATAG 642
Db 511 TTTCCGGCGCGTTCAGTTCGGTGTCTGGGCTCTGAGGTGGGTCTGGGATAG 570
QY 643 GTTCTCGCGGGTCTGATGTGCGCGCGCTTCCGCTATGTGGCTGATGAGCGTCA 702
Db 571 GTTCTCGCGGGTCTGATGTGCGCGCGCTTCCGCTATGTGGCTGATGAGCGTCA 630
QY 703 CCGCGGGCAGCGCGAGTCAACCGCGCGCTGAGTTCGGGTCTGCGCGCGCTTACGAGA 762
Db 631 CCGCGGGCAGCGCGAGTCAACCGCGCGCTGAGTTCGGGTCTGCGCGCGCTTACGAGA 690
QY 763 CGCGTATGGCTGACCGTGCCTCGCGCGGTGATCGCGAGAACCGTGTCTGAATGATGA 822
Db 691 CGCGTATGGCTGACCGTGCCTCGCGCGGTGATCGCGAGAACCGTGTCTGAATGATGA 750
QY 823 TTCTGATGAGCAACACTCTTTGGGGCAAAACACCCCGGATCGCGGTCAACGAGCGG 882
Db 751 TTCTGATGAGCAACACTCTTTGGGGCAAAACACCCCGGATCGCGGTCAACGAGCGG 810
QY 883 AATACGGCAGATGTGGGCGCAACAGCCCGCGGTGTTGGCTACGCCGCGCGAGCG 942
Db 811 AATACGGCAGATGTGGGCGCAACAGCCCGCGGTGTTGGCTACGCCGCGCGAGCG 870
QY 943 CGACGGCAGCGCGACGTTGCTGCTTCGAGAGGCGCGGAGATGACACGCGCGGTG 1002
Db 871 CGACGGCAGCGCGACGTTGCTGCTTCGAGAGGCGCGGAGATGACACGCGCGGTG 930
QY 1003 GGCTCTCAGCAGCGCGCGGTGAGAGGCGCTCGACACCGCGCGCGGAGACCACT 1062
Db 931 GGCTCTCAGCAGCGCGCGGTGAGAGGCGCTCGACACCGCGCGCGGAGACCACT 990
QY 1063 TGATGAACATGTGCCCGAGGCGTCAACAGCTGGGCGCGCGCGCGCGCGCGCGCG 1122
Db 991 TGATGAACATGTGCCCGAGGCGTCAACAGCTGGGCGCGCGCGCGCGCGCGCGCG 1050
QY 1123 CTTCTTCAAGCTGGGTGGCTGTGGAAGACGCTCTCGCGCGCGCGCGCGCGCGCGCG 1182
Db 1051 CTTCTTCAAGCTGGGTGGCTGTGGAAGACGCTCTCGCGCGCGCGCGCGCGCGCGCG 1110
QY 1243 ACACCTTGAATGATGTTGAGGGCTTTGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1302
Db 1171 ACACCTTGAATGATGTTGAGGGCTTTGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1230
QY 1303 CCGCGCGCAAAACGCGGTTCGCGCGATGAGTCTGCTGCGCGGCGCGCGCGCTGCGG 1362
Db 1231 CCGCGCGCAAAACGCGGTTCGCGCGATGAGTCTGCTGCGCGGCGCGCGCGCTGCGG 1290
QY 1363 GTCTGGCGGTGGGTGGCGCGCACTTGGGTGGCGCGCGCGCGCTGGTCTGGTCTGG 1422
Db 1291 GTCTGGCGGTGGGTGGCGCGCACTTGGGTGGCGCGCGCGCGCTGGTCTGGTCTGG 1350
QY 1423 TGCGCGAGCGCTGGCGCGCGCAACACGAGCAGTCAACCGCGCGCGCGCGCGCTGCGCG 1482
Db 1351 TGCGCGAGCGCTGGCGCGCGCAACACGAGCAGTCAACCGCGCGCGCGCGCGCTGCGCG 1410
QY 1483 TGACAGCTGACAGCGCGCGGAAAGAGGCGCGCGCGAGATGCTGGCGCGCGCTGCGCG 1542
Db 1411 TGACAGCTGACAGCGCGCGGAAAGAGGCGCGCGCGAGATGCTGGCGCGCGCTGCGCG 1470
QY 1543 TGGGCGAGATGGCGCGCGGTGGTGGGTCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 1602
Db 1471 TGGGCGAGATGGCGCGCGGTGGTGGGTCAAGTGTGCTGCTGCTGCTGCTGCTGCTG 1530
QY 1603 GACCTATGATGCGCGATTCCTCGGCGAGCGGCGA 1639
Db 1531 GACCTATGATGCGCGATTCCTCGGCGAGCGGCGTA 1567

QY 1363 GTCTGGCGGTGGGTGCGCCGCCAATCTGGGTGGGCGGCGCTCGGTGGTTCGTTGTCGG 1422
 Db 1291 GTCTGGCGGTGGGTGCGCCGCCAATCTGGGTGGGCGGCGCTCGGTGGTTCGTTGTCGG 1350
 QY 1423 TGCCGCGAGCTGGGCGCGCGCCCAACAGGAGTCACTCCCGCGCGCGCGCGCTGCGCGC 1482
 Db 1351 TGCCGCGAGCTGGGCGCGCGCCCAACAGGAGTCACTCCCGCGCGCGCGCGCTGCGCGC 1410
 QY 1483 TGACAGGCTGACAGAGCCCGCGGAAAGAGGCGCGCGAGATGCTGGGCGGCGTTCGCGG 1542
 Db 1411 TGACAGGCTGACAGAGCCCGCGGAAAGAGGCGCGCGAGATGCTGGGCGGCGTTCGCGG 1470
 QY 1543 TGGGCGAGATGGGCGCGAGGCGCGGTGGTGGGCTCAGTGGTGTCTCGGTTCGCGCGC 1602
 Db 1471 TGGGCGAGATGGGCGCGAGGCGCGGTGGTGGGCTCAGTGGTGTCTCGGTTCGCGCGC 1530
 QY 1603 GACCTATGTGATGCGCGCATCTCCGCGAGCGGCGA 1639
 Db 1531 GACCTATGTGATGCGCGCATCTCCGCGAGCGGCGTA 1567

RESULT 14
 AR353355
 LOCUS
 DEFINITION Sequence 106 from patent US 6592877.
 ACCESSION AR353355
 VERSION AR353355.1 GI:33759161
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 3058)
 AUTHORS Read,S.G., Skeiky,Y.A.W., Dillon,D.C., Campos-Neto,A., Houghton,R.,
 Vedvick,T.S., Twardzik,D.R., Lodes,M.J. and Hendrickson,R.C.
 TITLE Compounds and methods for immunotherapy and diagnosis of
 tuberculosis
 JOURNAL Patent: US 6592877-A 106 15-JUL-2003;
 FEATURES source Location/Qualifiers
 1..3058
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Query Match 51.1%; Score 1169; DB 6; Length 3058;
 Best Local Similarity 99.6%; Pred. No. 2,7e-135;
 Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 463 TCATGTTGGATTTCGGGCGCTTACCACCGGAGATCAACTCCGCGAGATGTCGCGCGCC 522
 Db 391 TAAATGGTGGATTTCGGGCGCTTACCACCGGAGATCAACTCCGCGAGATGTCGCGCGCC 450
 QY 523 CGGGTTTCGGCTCGCTGGTGGCGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGAAGTGT 582
 Db 451 CGGGTTTCGGCTCGCTGGTGGCGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGAAGTGT 510
 QY 583 TTTCGGCGCGGTGGCGCTTTCAGTGGTGGTTCGCGGTCTGAGCGTGGGTCGTTGATGAG 642
 Db 511 TTTCGGCGCGGTGGCGCTTTCAGTGGTGGTTCGCGGTCTGAGCGTGGGTCGTTGATGAG 570
 QY 643 GTTCGTTCGGCGGTCTGATGTGGCGCGCGCTCGCGGTATGTGGCGTGGATGAGCGTCA 702
 Db 571 GTTCGTTCGGCGGTCTGATGTGGCGCGCGCTCGCGGTATGTGGCGTGGATGAGCGTCA 630
 QY 703 CCGCGGCGAGCGCGAGCTGACCGCGCGCGCGCGCGGTGTCGCGGCGCGCTACGAGA 762
 Db 631 CCGCGGCGAGCGCGAGCTGACCGCGCGCGCGCGCGGTGTCGCGGCGCGCTACGAGA 690
 QY 763 CGGGTATGGGTGACCGTGCCTCCCGCGCGGTGATCGCGAGAACCGTGTGTAAGTATGA 822
 Db 691 CGGGTATGGGTGACCGTGCCTCCCGCGCGGTGATCGCGAGAACCGTGTGTAAGTATGA 750
 QY 823 TTTCGATAGGCAACCTTTGGGCGAAACACCGCGCGATCGCGGTCAACGAGCGCG 882

Db 751 TTTCGATAGCGAACCACTCTTGGGGGAAAACACCCCGCGCATCGCGGTCAACGAGGCGG 810
 QY 883 AATACGGGAGATGTGGGCCCCAAGACGCGCGCGGATCTTTGGCTACGCCCGCGGACGG 942
 Db 811 AATACGGGAGATGTGGGCCCCAAGACGCGCGCGGATCTTTGGCTACGCCCGCGGACGG 870
 QY 943 CGACGGCGAGCGGAGCTGTGCTCGGTTTCGAGGAGGCGCGGAGATGACACGAGCGGGTG 1002
 Db 871 CGACGGCGAGCGGAGCTGTGCTCGGTTTCGAGGAGGCGCGGAGATGACACGAGCGGGTG 930
 QY 1003 GGCTCTCTCGAGCAGCGCGCGGTTCGAGGAGGCGCTCCGACACCGCGCGCGCAACCACT 1062
 Db 931 GGCTCTCTCGAGCAGCGCGCGGTTCGAGGAGGCGCTCCGACACCGCGCGCGCAACCACT 990
 QY 1063 TGATGAACAATGTGCGCGCGAGGCGCTGCAACAGCTGGCCAGCCACCGAGGCGACCGC 1122
 Db 991 TGATGAACAATGTGCGCGCGAGGCGCTGCAACAGCTGGCCAGCCACCGAGGCGACCGC 1050
 QY 1123 CTTCCTTCAAGCTGGGTGGCTGTGGAAGACGGTCTCGCCGCGATCGGTTCGCGCATCAGCA 1182
 Db 1051 CTTCCTTCAAGCTGGGTGGCTGTGGAAGACGGTCTCGCCGCGATCGGTTCGCGCATCAGCA 1110
 QY 1183 ACATGGTTCGATGGCCCAACCAACCAATGTTCGATGACCAACTCGGGTGTTCGATGACCA 1242
 Db 1111 ACATGGTTCGATGGCCCAACCAACCAATGTTCGATGACCAACTCGGGTGTTCGATGACCA 1170
 QY 1243 ACACCTTCAGCTTCGATGTTGAAGGCTTTGCTCCGCGCGCGCGCGCGCGCGTGCAGAA 1302
 Db 1171 ACACCTTCAGCTTCGATGTTGAAGGCTTTGCTCCGCGCGCGCGCGCGCGCGTGCAGAA 1230
 QY 1303 CGCGCGCGCAAAACCGGGTCCGCGCGATGAGCTCGCTGGGCGAGTCTCGTGGGTCTTCGG 1362
 Db 1231 CGCGCGCGCAAAACCGGGTCCGCGCGATGAGCTCGCTGGGCGAGTCTCGTGGGTCTTCGG 1290
 QY 1363 GTCTGGGCGGTGGGTGCGCGCGCAACTTGGGTTCGGGCGGCTCGGTTCGTTTCGG 1422
 Db 1291 GTCTGGGCGGTGGGTGCGCGCGCAACTTGGGTTCGGGCGGCTCGGTTCGTTTCGG 1350
 QY 1423 TGCCGCGAGCTGGGCGCGCGCGCAACAGGAGTCACTCCCGCGCGCGCGCGCGTGCAGC 1482
 Db 1351 TGCCGCGAGCTGGGCGCGCGCGCAACAGGAGTCACTCCCGCGCGCGCGCGCGTGCAGC 1410
 QY 1483 TCACAGCTGACACAGCGCGCGGAAAGAGGCGCGCGGAGATGCTGGGCGGCTGCGCG 1542
 Db 1411 TGACAGCTGACACAGCGCGCGGAAAGAGGCGCGCGGAGATGCTGGGCGGCTGCGCG 1470
 QY 1543 TGGGCGAGATGGGCGCGCGCGCGCGCGCGTGGTGGGCTCAGTGGTGTCTCGGTTCGCGCG 1602
 Db 1471 TGGGCGAGATGGGCGCGCGCGCGCGCGCGTGGTGGGCTCAGTGGTGTCTCGGTTCGCGCG 1530
 QY 1603 GACCTATGTGATGCGCGCATCTCCGCGAGCGGCGA 1639
 Db 1531 GACCTATGTGATGCGCGCATCTCCGCGAGCGGCGTA 1567

RESULT 15
 AX429698
 LOCUS
 DEFINITION Sequence 106 from Patent EP1203817.
 ACCESSION AX429698
 VERSION AX429698.1 GI:21540898
 KEYWORDS
 SOURCE unidentified
 ORGANISM unidentified
 unclassified.
 REFERENCE 1
 AUTHORS Read,S.G., Skeiky,Y.A., Dillon,D.C., Campos-Neto,A., Houghton,R.L.,
 Vedvick,T.S. and Twardzik,D.R.
 TITLE Compounds and methods for immunotherapy and diagnosis of
 tuberculosis
 JOURNAL Patent: EP 1203817-A 106 08-MAY-2002;
 FEATURES Location/Qualifiers


```
source      1. .3058
            /organism="unidentified"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32644"

ORIGIN
Query Match      51.1%; Score 1169; DB 6; Length 3058;
Best Local Similarity 99.6%; Pred. No. 2.7e-135;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 463 TCATGTGATTTTCGGGGGCTTACCAACCGAGATCAACTCCGCGAGAGATGTACGCCGCC 522
Db 391 TAATGTGATTTTCGGGGGCTTACCAACCGAGATCAACTCCGCGAGAGATGTACGCCGCC 450
Qy 523 CGGGTTCCGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGCTGGCGAGTGAACCTGT 582
Db 451 CGGGTTCCGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGCTGGCGAGTGAACCTGT 510
Qy 583 TTTCCGCGCGCTCGGCGTTTCACTCGTGGTCTTGGGGTCTGACGGTGGGGTCTGGATAG 642
Db 511 TTTCCGCGCGCTCGGCGTTTCACTCGTGGTCTTGGGGTCTGACGGTGGGGTCTGGATAG 570
Qy 643 GTTCTGCGCGGCTCTGATGGTGGCGCGGCTCTCGCGTATGTGGGTGTGATCAGCGTCA 702
Db 571 GTTCTGCGCGGCTCTGATGGTGGCGCGGCTCTCGCGTATGTGGGTGTGATCAGCGTCA 630
Qy 703 CCGCGGCGAGCGCGAGCTGACCGCGCGCCAGGTCCGGTTGCTGGCGGGCTACGACA 762
Db 631 CCGCGGCGAGCGCGAGCTGACCGCGCGCCAGGTCCGGTTGCTGGCGGGCTACGACA 690
Qy 763 CGCGGTATGGGCTGACGGTGGCGCGCGCGGTGATCCCGAGAACCGTGTGTAACGTATGA 822
Db 691 CGCGGTATGGGCTGACGGTGGCGCGCGCGGTGATCCCGAGAACCGTGTGTAACGTATGA 750
Qy 823 TTCTGTAGCGACCAACCTCTTGGGGCAAAACACCCCGGATCGCGGTCAACGAGGCGG 882
Db 751 TTCTGTAGCGACCAACCTCTTGGGGCAAAACACCCCGGATCGCGGTCAACGAGGCGG 810
Qy 883 AATAACGCGAGATGTGGGCGCAAGACCGCGCGGATGTTTGGCTACGCGCGGCGAAGCG 942
Db 811 AATAACGCGAGATGTGGGCGCAAGACCGCGCGGATGTTTGGCTACGCGCGGCGAAGCG 870
Qy 943 CGACGGGACGCGGAGCTTGTCTCGCTTCCAGAGGCGCGGAGATGACAGCGGGGTG 1002
Db 871 CGACGGGACGCGGAGCTTGTCTCGCTTCCAGAGGCGCGGAGATGACAGCGGGGTG 930
Qy 1003 GGCTCTTCGAGCAGGCGCGCGCGGTGAGGAGGCTCCGACACCGCGCGGCGAACCAGT 1062
Db 931 GGCTCTTCGAGCAGGCGCGCGGTGAGGAGGCTCCGACACCGCGCGGCGAACCAGT 990
Qy 1063 TGATGAACAATGTGCTCCAGGCGCTGCAACAGCTGGCCCHAGCCACGAGGCGACCCAGC 1122
Db 991 TGATGAACAATGTGCTCCAGGCGCTGCAACAGCTGGCCCHAGCCACGAGGCGACCCAGC 1050
Qy 1123 CTTCCTCCAAAGCTGGGTGGCTGTGGAAGACGCTCTCGCGCATCGGTCCCGATCAGCA 1182
Db 1051 CTTCCTCCAAAGCTGGGTGGCTGTGGAAGACGCTCTCGCGCATCGGTCCCGATCAGCA 1110
Qy 1183 ACATGGTGTGATGGCGCAACCAACATGTGATGACCAACTCGGGTGTGATGACCA 1242
Db 1111 ACATGGTGTGATGGCGCAACCAACATGTGATGACCAACTCGGGTGTGATGACCA 1170
Qy 1243 ACACCTTGAGCTGATGTTGAAGGCTTTCCTCGCGGGCGCGCGCGCGCGCGTGCMAA 1302
Db 1171 ACACCTTGAGCTGATGTTGAAGGCTTTCCTCGCGGGCGCGCGCGCGCGCGTGCMAA 1230
Qy 1303 CCGCGGCGCAAAACCGGGTCCGGCGATGAGCTCGTGGGCGAGCTCGTGGGTTCCTCGG 1362
Db 1231 CCGCGGCGCAAAACCGGGTCCGGCGATGAGCTCGTGGGCGAGCTCGTGGGTTCCTCGG 1290
Qy 1363 GTCTGGCGGTGGGGTGGCGCCAACTTGGGTGGCGGGCGCTCGTGGGTTCCTCGG 1422
Db 1291 GTCTGGCGGTGGGGTGGCGCCAACTTGGGTGGCGGGCGCTCGTGGGTTCCTCGG 1350
```

Search completed: July 3, 2004, 20:19:30
Job time : 8840 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2004, 14:04:57 ; Search time 899 Seconds
(without alignments)
10807.151 Million cell updates/sec

Title: US-09-597-796c-11
Perfect score: 2287
Sequence: 1 tetagaataattttgttta.....ggntgtaacaagccgaaa 2287

Scoring table: IDENTITY.NUC
Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2284	99.9	2287	6	AAD47083 Mycobacte
2	2284	99.9	2287	6	AAD28342 Mycobacte
3	2266.4	99.1	2287	2	Aaz20194 Mycobacte
4	2256.2	98.7	2286	6	Abk14128 DNA encod
5	2187.4	95.6	2191	4	AAL40773 Nucleotid
6	2186.8	95.6	2451	8	Ada26360 Mycobacte
7	2186.8	95.6	2457	8	Ada26359 Mycobacte
8	2186.8	95.6	2637	8	Ada26358 Mycobacte
9	2186.8	95.6	2808	6	AAD47110 Mycobacte
10	2186.8	95.6	2808	8	Ada26357 Mycobacte
11	2186.8	95.6	3060	8	Ada26363 M. bovis
12	2186.8	95.6	3104	8	Ada26362 Mycobacte
13	2186.8	95.6	3474	8	Ada26361 Mycobacte
14	2185.2	95.5	2190	6	AAD47084 Mycobacte
15	2185.2	95.5	2190	6	AAD28343 Mycobacte
16	1770.8	77.4	1797	2	Aaz20205 Mycobacte
17	1770.8	77.4	1797	6	AAD47086 Mycobacte
18	1770.8	77.4	1797	6	AAD28344 Mycobacte
19	1770.8	77.4	1797	6	Abk14139 DNA encod
20	1582	69.2	3030	8	Ada26355 Mycobacte
21	1572.6	68.8	2181	8	Ada26353 Mycobacte
22	1183.2	51.7	2445	5	AAS03792 M. tuberc
23	1175.4	51.4	2365	5	AAS03791 M. tuberc

24	1175.2	51.4	2232	5	AAS03790 M. tuberc
25	1172.8	51.3	1801	2	Aaz20203 Mycobacte
26	1169	51.1	3058	2	Aav44395 Mycobacte
27	1169	51.1	3058	2	Aav64503 M. tuberc
28	1169	51.1	3058	2	Aaz19093 M. tuberc
29	1169	51.1	3058	2	Aaz19093 M. tuberc
30	1169	51.1	3058	5	AAS03779 M. tuberc
31	1169	51.1	3058	6	AAD47082 Mycobacte
32	1169	51.1	3058	6	AAD28341 Mycobacte
33	1169	51.1	75216	6	ABX09141 Mycobacte
34	1169	51.1	110000	4	Continuation (14 o
35	1169	51.1	110000	4	Continuation (14 o
36	1168.6	51.1	1176	2	AAX34030 Mycobacte
37	1167.4	51.0	1188	2	AAX34031 Mycobacte
38	1164.2	50.9	1565	2	AAT91455 Mycobacte
39	1164.2	50.9	1565	2	AAT91521 Mycobacte
40	1160.8	50.8	1800	6	ABK14137 DNA encod
41	955	41.8	110000	4	Continuation (39 o
42	947.4	41.4	110000	4	Continuation (16 o
43	942.6	41.2	110000	4	Continuation (16 o
44	941	41.1	3027	2	Aav44397 Mycobacte
45	941	41.1	3027	2	Aav64506 M. tuberc

ALIGNMENTS

RESULT 1
AAD47083
ID AAD47083 standard; DNA; 2287 BP.
XX AAD47083;
AC AAD47083;
XX 29-AUG-2003 (revised)
DT 27-JAN-2003 (first entry)
XX Mycobacterium sp. MTB72F fusion protein encoding DNA.
XX Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen; Ra35;
XX Ra12; MTB72F; chimeric; gene; ds.
XX Mycobacterium sp.
OS Mycobacterium tuberculosis.
OS Chimeric.

XX Key Location/Qualifiers
FT CDS 42..2231
FT /*tag= a
FT /product= "MTB72F fusion protein"

XX WO200272792-A2.
XX 19-SEP-2002.
XX 13-MAR-2002; 2002WO-US008223.
XX 13-MAR-2001; 2001US-0275837P.
XX (CORI-) CORIYA CORP.
XX Skeiky Y, Brannon M, Guderian J;
XX WPI; 2002-759844/82.
XX P-PSDB; AAE29708.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.
XX Disclosure; Page 87-90; 155pp; English.
XX The invention relates to a recombinant nucleic acid molecule encoding a

CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from T8A, Leif, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is a
 CC DNA encoding MTB72F fusion protein. This fusion protein comprises Ral2
 CC and Ra35 protein from Mycobacterium tuberculosis and Tbh9 protein from
 CC Mycobacterium sp. (Updated on 29-AUG-2003 to standardise OS field)
 XX
 SQ Sequence 2287 BP; 372 A; 717 C; 814 G; 381 T; 0 U; 3 Other;

Query Match 99.9%; Score 2284; DB 6; Length 2287;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TCTAGAAATATTTTGTGTTTACTTTAAGAANGANATATACATATGCATCACCATC	60
DB	1	TCTAGAAATATTTTGTGTTTACTTTTAAAGAANGANATATACATATGCATCACCATC	60
QY	61	ACAGGCGCGTCCGATAAATCTCCAGCTGTCCAGGCTGGCGAGGATTGCGCATTC	120
DB	61	ACAGGCGCGTCCGATAAATCTCCAGCTGTCCAGGCTGGCGAGGATTGCGCATTC	120
QY	121	TCGGGCGAGCGATGGCGATCGCGGCGAGATCCGATCGGTTGGGGGTACCCACCGTTC	180
DB	121	TCGGGCGAGCGATGGCGATCGCGGCGAGATCCGATCGGTTGGGGGTACCCACCGTTC	180
QY	181	ATATCGGGCTTACCGCTTCTCGCTTGGGTGTTGTCGACAAACGCGAAACGCGGCAC	240
DB	181	ATATCGGGCTTACCGCTTCTCGCTTGGGTGTTGTCGACAAACGCGAAACGCGGCAC	240
QY	241	GAGTCCAAACGCGTGGTGGGAGCGCTCCGGCGGCAAGTCTCGGCATCTCCACCGGCGAGC	300
DB	241	GAGTCCAAACGCGTGGTGGGAGCGCTCCGGCGGCAAGTCTCGGCATCTCCACCGGCGAGC	300
QY	301	TGATCACCGGCTGACGGGCTCCGATCAATCTGGCCACCGGATGGCGGACCGGCTTA	360
DB	301	TGATCACCGGCTGACGGGCTCCGATCAATCTGGCCACCGGATGGCGGACCGGCTTA	360
QY	361	ACGGGCTATCATCCGGTGAAGTCTATCTCGGTGACCTGGCAACCAAGTGGGCGGACGC	420
DB	361	ACGGGCTATCATCCGGTGAAGTCTATCTCGGTGACCTGGCAACCAAGTGGGCGGACGC	420
QY	421	GTCAGGGAAAGTGAATTCGGCGAGGAAACCCCGGCGGAATTCATGGTGGATTTCCGGG	480
DB	421	GTCAGGGAAAGTGAATTCGGCGAGGAAACCCCGGCGGAATTCATGGTGGATTTCCGGG	480
QY	481	CGTTTACCAACCGGAGATCAATCTCCGCGAGGATGTACCGCGCCCGGCTTGGCTCTCG	540
DB	481	CGTTTACCAACCGGAGATCAATCTCCGCGAGGATGTACCGCGCCCGGCTTGGCTCTCG	540
QY	541	TGGCCGCGCTCAGATGTGGAAACAGCGTGGCGAGTGACCTGTTTCGGCCGGTCTGGCGT	600
DB	541	TGGCCGCGCTCAGATGTGGAAACAGCGTGGCGAGTGACCTGTTTCGGCCGGTCTGGCGT	600
QY	601	TTCACTCGGTGGTCTGGGGTCTACGCTGGGGTCTGGATAGTTCGTTGGGGGTCTGA	660
DB	601	TTCACTCGGTGGTCTGGGGTCTACGCTGGGGTCTGGATAGTTCGTTGGGGGTCTGA	660
QY	661	TGTTGGCGGCGCTCGCCGTATGTGGCGTGGATGAGCGTCAACCGGGCGGCGAGCGAGC	720
DB	661	TGTTGGCGGCGCTCGCCGTATGTGGCGTGGATGAGCGTCAACCGGGCGGCGAGCGAGC	720
QY	721	TGACCGCGCGGCTCGGGTCTGGCGGCGGCTACGAGACGGCGTATGGGCTGACGG	780
DB	721	TGACCGCGCGGCTCGGGTCTGGCGGCGGCTACGAGACGGCGTATGGGCTGACGG	780

QY	781	TGCCCCCGCGGTGATTCGCCGAGAACCGTGTGAACCTGATGATTCTGATAGCAACACC	840
DB	781	TGCCCCCGCGGTGATTCGCCGAGAACCGTGTGAACCTGATGATTCTGATAGCAACACC	840
QY	841	TCTTGGGGCAAAACACCCCGCGGATCGCGGTCAACAGAGCCGCAATACCGGAGATGTGGG	900
DB	841	TCTTGGGGCAAAACACCCCGCGGATCGCGGTCAACAGAGCCGCAATACCGGAGATGTGGG	900
QY	901	CCCAAGACGCCCGCGATGTTTGGCTACGCGCGGCGCACCGGCGAGCGGCGAGCT	960
DB	901	CCCAAGACGCCCGCGATGTTTGGCTACGCGCGGCGCACCGGCGAGCGGCGAGCT	960
QY	961	TGCTGCGCTTCAGAGAGGCGCGGAGATGACAGCGCGGGTGGGCTCTCTCAGACAGGCG	1020
DB	961	TGCTGCGCTTCAGAGAGGCGCGGAGATGACAGCGCGGGTGGGCTCTCTCAGACAGGCG	1020
QY	1021	CCGCGGTTCAGGAGGCTCCGACACCGCGCGGCGAAACAGTTGATGAACATGTGCCCC	1080
DB	1021	CCGCGGTTCAGGAGGCTCCGACACCGCGCGGCGAAACAGTTGATGAACATGTGCCCC	1080
QY	1081	AGGCGCTGCAACAGCTGGCCGACGCCACGAGGGCACCAACGCTTCTTCCAAAGTGGGTG	1140
DB	1081	AGGCGCTGCAACAGCTGGCCGACGCCACGAGGGCACCAACGCTTCTTCCAAAGTGGGTG	1140
QY	1141	GCCTGTGAAGACGCTTCGCGGATCGGTCGCGATCAGCAACATGTCGATGCGCA	1200
DB	1141	GCCTGTGAAGACGCTTCGCGGATCGGTCGCGATCAGCAACATGTCGATGCGCA	1200
QY	1201	ACAACCAATGTCGATGACCAACTCGGCTGTGTCGATGACCAACCTTGAGCTCGATGT	1260
DB	1201	ACAACCAATGTCGATGACCAACTCGGCTGTGTCGATGACCAACCTTGAGCTCGATGT	1260
QY	1261	TGAAGGGCTTGTCTCCGCGGCGCGCCGAGGCGGTGCAACCGCGCGCAACCGGG	1320
DB	1261	TGAAGGGCTTGTCTCCGCGGCGCGCCGAGGCGGTGCAACCGCGCGCAACCGGG	1320
QY	1321	TCCGGCGCATGAGCTCGCTGGGCGAGCTCGCTGGGTTCTTTCGGGCTCGGGGTGG	1380
DB	1321	TCCGGCGCATGAGCTCGCTGGGCGAGCTCGCTGGGTTCTTTCGGGCTCGGGGTGG	1380
QY	1381	CGGCCAATTCGGGTTCGGCGGCGCTCGGTCGCTTCTTGTTCGGTCCGAGGCTGGGCGG	1440
DB	1381	CGGCCAATTCGGGTTCGGCGGCGCTCGGTCGCTTCTTGTTCGGTCCGAGGCTGGGCGG	1440
QY	1441	CGGCCAACAGGAGTCAACCCCGCGCGCGGCGGTGCGCTGACCAAGCTGACCAAGCG	1500
DB	1441	CGGCCAACAGGAGTCAACCCCGCGCGCGGCGGTGCGCTGACCAAGCTGACCAAGCG	1500
QY	1501	CGCGGAAAGAGGCGCGGCGAGATGTGGGCGGCTGCGGTTGGGCGAGATGGGCGCA	1560
DB	1501	CGCGGAAAGAGGCGCGGCGAGATGTGGGCGGCTGCGGTTGGGCGAGATGGGCGCA	1560
QY	1561	GGGCGGTGGTGGGCTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1620
DB	1561	GGGCGGTGGTGGGCTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1620
QY	1621	ATTCTCCGGCAGCGCGGATATCCCGCGCGGCTTGTTCGAGGACCGGTTTCGCGACT	1680
DB	1621	ATTCTCCGGCAGCGCGGATATCCCGCGCGGCTTGTTCGAGGACCGGTTTCGCGACT	1680
QY	1681	TCCCGGCTGCGCTCGACCCGCTCGGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1740
DB	1681	TCCCGGCTGCGCTCGACCCGCTCGGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1740
QY	1741	ACATCAACACCAACTGGGTTTCAACCAACCGCGTGGGCGCGGCGGCGGCGGCGGCGG	1800
DB	1741	ACATCAACACCAACTGGGTTTCAACCAACCGCGTGGGCGCGGCGGCGGCGGCGGCGG	1800
QY	1801	ATCCCAACGCTGCTGCTGACCAACCAACCGCTGATCGCGGCGGCGGCGGCGGCGGCGG	1860
DB	1801	ATCCCAACGCTGCTGCTGACCAACCAACCGCTGATCGCGGCGGCGGCGGCGGCGGCGG	1860
QY	1861	CGTTTCAGCGTCCGCTCCGCGCAACCTTACGCGGTCGATGCTGGGTATGACCGCACCC	1920

Db	1861	CGTTACGGTCGGCTCCGGCCAAACCTACGGCGTCGATGTGTCGGGTATGACCGCACCC	1920
Qy	1921	AGATGTGGGTGCTGCTACGCTGGCGGTGGCGTGCCTTGGCTGGCGGCGCATCGGTG	1980
Db	1921	AGGATGTGGGTGCTGCTACGCTGGCGGTGGCGTGCCTTGGCTGGCGGCGCATCGGTG	1980
Qy	1981	GCSCGTGCGGGTTGGTGAGCCCGTCTGCGCATGGGCAACAGCGGTGGCAGGGCGGAA	2040
Db	1981	CGCGGTGCGGGTTGGTGAGCCCGTCTGCGCATGGGCAACAGCGGTGGCAGGGCGGAA	2040
Qy	2041	CGCCCCGTGCGGTGCTGGCAGGGTGTCGCGTTCGCCAAACCGTGCAGGGCTCGGATT	2100
Db	2041	CGCCCCGTGCGGTGCTGGCAGGGTGTCGCGTTCGCCAAACCGTGCAGGGCTCGGATT	2100
Qy	2101	CGTGCACCGTGGCAGAGACATTGAACGGGTTCATCCAGTTCGATGCCCGCATCCAGC	2160
Db	2101	CGTGCACCGTGGCAGAGACATTGAACGGGTTCATCCAGTTCGATGCCCGCATCCAGC	2160
Qy	2161	CCGCTGATTCCGGCGGGCCCGTCTCTCAACGGCCTPAGCACAGGTGGTCGGTATGAACACGG	2220
Db	2161	CCGCTGATTCCGGCGGGCCCGTCTCTCAACGGCCTPAGCACAGGTGGTCGGTATGAACACGG	2220
Qy	2221	CCGCGTCTTAGGATATCCATCACACTGGCGGCGCTCGACAGATCCGNGTGAACAAG	2280
Db	2221	CCGCGTCTTAGGATATCCATCACACTGGCGGCGCTCGACAGATCCGNGTGAACAAG	2280
Qy	2281	CCCGAAA	2287
Db	2281	CCCGAAA	2287

RESULT 2
AAD28342
ID AAD28342 standard; DNA; 2287 BP.
XX
XX
AAD28342;
XX
XX
22-APR-2002 (first entry)
XX
XX
Mycobacterium species MTB72F fusion protein encoding DNA.
DE
XX
XX
Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB72F; Ral2-TbH9-Ra35; ds.
KW

XX	WO200198460-A2.
FN	
XX	
XX	27-DEC-2001.
PD	
XX	
XX	20-JUN-2001; 2001WO-US019959.
PF	
XX	
XX	20-JUN-2000; 2000US-00597796.
PR	
PR	01-FEB-2001; 2001US-028537P.
XX	
XX	(CORI-) CORIXA CORP.
PA	
XX	
PI	Skeiky Y, Reed S, Alderson M;

WPI; 2002-147798/19.
P-PSDB; AAE17572.

Composition comprising MTB39 antigen and MTB32A antigen from
Mycobacterium species, useful for eliciting immune response in a subject.

Claim 62; Page 103-106; 136pp; English.

The present invention relates to fusion proteins containing at least two
Mycobacterium species antigens, nucleotides encoding them and
compositions comprising such fusion proteins. The present invention
particularly relates to nucleic acids encoding fusion proteins that
include two or more individual M. tuberculosis antigens which increase
the serological sensitivity of sera from individuals infected with
tuberculosis and methods for their use in diagnosis, prevention and
treatment of tuberculosis infection. Sequences of the invention are
useful for eliciting an immune response in a mammal, e.g., human,
immunised with BCG. They are useful in the diagnosis, treatment and
prevention of Mycobacterium infection. The fusion proteins and the
polynucleotides are useful as diagnostic tools in patients infected with
Mycobacterium, in vitro and in vivo assays for detecting humoral
antibodies or cell-mediated immunity against M. tuberculosis, for the
diagnosis of an infection or monitoring of disease progression, as
immunogens to generate or elicit a protective immune response in a
patient and for raising anti-M. tuberculosis antibodies in a non-human
animal. Sequences of the invention are also used as vaccines. MTB32A
fusion proteins of the invention are useful as in vivo diagnostic agents
for intradermal skin test. The present sequence is a DNA encoding
Mycobacterium species MTB72F (Ra12-TbH9-Ra35) fusion protein

Sequence 2287 BP; 372 A; 717 C; 814 G; 381 T; 0 U; 3 Other;

541 TGGCCGCGCTCAGATGAGGACAGCGTGGAGTACCTGTTTTCGCGCGCGTGGCGT 600
541 TGGCCGCGCTCAGATGAGGACAGCGTGGAGTACCTGTTTTCGCGCGCGTGGCGT 600
601 TTGAGTCGGTGGTCTGAGGCTGAGCGTGGGCTGCTGATAGTTTCTGCGCGGCTGGA 660
601 TTGAGTCGGTGGTCTGAGGCTGAGCGTGGGCTGCTGATAGTTTCTGCGCGGCTGGA 660
661 TGGTGGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 720
661 TGGTGGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 720
721 TGAACCGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 780
721 TGAACCGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 780
781 TGGCCCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 840
781 TGGCCCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 840
841 TCTTGGGCAAAACACCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 900
841 TCTTGGGCAAAACACCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 900
901 CCGAGAGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 960
901 CCGAGAGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 960
961 TGGTGGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1020
961 TGGTGGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1020
1021 CCGCGCTGAGGAGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1080
1021 CCGCGCTGAGGAGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1080
1081 AGGCGCTGAGGAGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1140
1081 AGGCGCTGAGGAGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1140
1141 GCCTGTGGAAGAGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1200
1141 GCCTGTGGAAGAGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1200
1201 ACACACATGTCGATGACCACTCGGCTGTCGATGACCACTGTCGATGACCACTGTCGATG 1260
1201 ACACACATGTCGATGACCACTCGGCTGTCGATGACCACTGTCGATGACCACTGTCGATG 1260
1261 TGAAGGCTTTGCTCGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1320
1261 TGAAGGCTTTGCTCGCGCGGCTCGCGCTATGAGCGTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 1320
1321 TCGCGCGATGAGCTCGCTGGGAGTCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1380
1321 TCGCGCGATGAGCTCGCTGGGAGTCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1380
1381 CCGCCAACTTGGGTGGGCGGCTCGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1440
1381 CCGCCAACTTGGGTGGGCGGCTCGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1440
1441 CCGCCAACTTGGGTGGGCGGCTCGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1500
1441 CCGCCAACTTGGGTGGGCGGCTCGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1500
1501 CCGCGAAAGAGGCGCGGCTCGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1560
1501 CCGCGAAAGAGGCGCGGCTCGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1560
1561 GGGCCGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1620
1561 GGGCCGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTG 1620
1621 ATTCTCGGCGAGCGGCGGATATCGCCCGCGGCTTGTGCGAGGACGCGTTCGCGGACT 1680

1621 ATTCTCGGCGAGCGGCGGATATCGCCCGCGGCTTGTGCGAGGACGCGTTCGCGGACT 1680
1681 TCCCGCGCTGCGCTCGACCGCTCGCGGATGCTGCCCAAGTGGGGCCACAGTGGTCA 1740
1681 TCCCGCGCTGCGCTCGACCGCTCGCGGATGCTGCCCAAGTGGGGCCACAGTGGTCA 1740
1741 ACATCAACACCAAACTGGGCTACCAACAGCGCTGGGCGCGGACCGGATGCTCATCG 1800
1741 ACATCAACACCAAACTGGGCTACCAACAGCGCTGGGCGCGGACCGGATGCTCATCG 1800
1801 ATCCCAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
1801 ATCCCAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
1861 GCTTCAGCGTGGGCTCGCGGCAACCTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
1861 GCTTCAGCGTGGGCTCGCGGCAACCTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
1921 AGGATGCTCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
1921 AGGATGCTCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
1981 GCGGCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
1981 GCGGCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
2041 GCGGCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
2041 GCGGCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
2101 GCTGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
2101 GCTGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
2161 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
2161 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
2221 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
2221 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
2281 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340
2281 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340

RESULT 3
AAZ20194
ID AAZ20194 standard; DNA; 2287 BP.
XX
AC AAZ20194;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mycobacterium tuberculosis antigen fusion protein Mtb32A DNA.
XX
KW Tuberculosis; antigen; fusion protein; Mtb32A; Ra12; TBH9; Ra35;
XX diagnosis; therapy; vaccine; immunogen; ss.
XX
OS Mycobacterium tuberculosis.
XX
PH Key Location/Qualifiers
CDS 42..2231
FT /*tag= a
XX
PN WO951748-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999;
XX 99WO-US007717.

DB 1741 GCCGGGACGGGATCGTCATCGATCCCAAGGTCGTGCTGCTGACCAACACCGTGCATC 1800
QY 1839 GCGGGGACCGACATCAATCGCTTACAGCTCGGCTCCGGCCAAACCTACGGGTCGAT 1898
DB 1801 GCGGGGACCGACATCAATCGCTTACAGCTCGGCTCCGGCCAAACCTACGGGTCGAT 1860
QY 1899 GTGGTCGGGTATGACCGACACAGGATGTCCGGTGTGTCAGCTGCGGGTCCGGTGGC 1958
DB 1861 GTGGTCGGGTATGACCGACACAGGATGTCCGGTGTGTCAGCTGCGGGTCCGGTGGC 1920
QY 1959 CTGCGCTCGGGCGGATCGGTGGGGCGGTCCGGTGTGTCAGCTGCGGGTCCGGTGGC 2018
DB 1921 CTGCGCTCGGGCGGATCGGTGGGGCGGTCCGGTGTGTCAGCTGCGGGTCCGGTGGC 1980
QY 2019 AACAGCGTGGCGAGGGCGGAACGCCCGCTCGGTGCTGCGAGGTCGCGCTCGGC 2078
DB 1981 AACAGCGTGGCGAGGGCGGAACGCCCGCTCGGTGCTGCGAGGTCGCGCTCGGC 2040
QY 2079 CAAACCGTGCAGGGTCGATTCCTGACCGGTCCGGAAGACATTTGAACGGGTTCATC 2138
DB 2041 CAAACCGTGCAGGGTCGATTCCTGACCGGTCCGGAAGACATTTGAACGGGTTCATC 2100
QY 2139 CAGTTCGATCGCGGATCCAGCCCGGTGATTCGGGGCGGCGCGCTCC 2228
DB 2101 CAGTTCGATCGCGGATCCAGCCCGGTGATTCGGGGCGGCGCGCTCC 2160
QY 2199 CAGGTGGTCGATGAACACACGCGCGCGCTCC 2228
DB 2161 CAGGTGGTCGATGAACACACGCGCGCGCTCC 2190

RESULT 8

ADA26358
ID ADA26358 standard; DNA; 2637 BP.
XX
AC ADA26358;
XX
DT 20-NOV-2003 (first entry)
XX
DE Mycobacterium MTB72F-Erd14 (fusion MTB89F) protein encoding DNA.
XX
DE ds; gene; fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
OS Chimeric.
OS Mycobacterium sp.
XX
XX
Key Location/Qualifiers
FT CDS 4..2631
FT /tag= a
FT /product= "MTB72F-Erd14 (fusion MTB89F)"
XX
PN WO2003070187-A2.
XX
XX
PD 28-AUG-2003.
XX
XX 18-FEB-2003; 2003WO-US004903.
XX
XX 15-FEB-2002; 2002US-0357351P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Guderian J, Reed S;
XX
XX WPI: 2003-697554/66.
XX P-PSDB; ADA26358.
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX

PS Claim 84; Fig 6; 112pp; English.

XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.

XX Sequence 2637 BP; 428 A; 840 C; 928 G; 441 T; 0 U; 0 Other;

Query Match 95.6%; Score 2186.8; DB 8; Length 2637;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 39 CATATGATACCATCACCATCACACGCGCGCTCCGATAACTTCCAGCTGTCCAGGGT 98

DB 1 CATATGATACCATCACCATCACACGCGCGCTCCGATAACTTCCAGCTGTCCAGGGT 60

QY 99 GGGCAGGATTCGCCATTCCGATCGGCGAGCGATGGCGATCGGGGCCAGATCCGATCG 158

DB 61 GGGCAGGATTCGCCATTCCGATCGGCGAGCGATGGCGATCGGGGCCAGATCCGATCG 120

QY 159 GGTGGGGGTCAACCCACCGTTTCATATCGGGGCTTACCGCTTCTCGGTTGGGTGTC 218

DB 121 GGTGGGGGTCAACCCACCGTTTCATATCGGGGCTTACCGCTTCTCGGTTGGGTGTC 180

QY 219 GACACACGCGACGCGCGAGTCCACCGCTGTCGGAGGCTCCGGCGGCAAGT 278

DB 181 GACACACGCGACGCGCGAGTCCACCGCTGTCGGAGGCTCCGGCGGCAAGT 240

QY 279 CTCGGCATCTCACCGCGAGCTGATCACCGCGTTCGACGGCGTCCGATCAACTCCGCC 338

DB 241 CTCGGCATCTCACCGCGAGCTGATCACCGCGTTCGACGGCGTCCGATCAACTCCGCC 300

QY 339 ACCCGATGGCGGACGCGCTTAACGGGATCATCCCGTTCGCTGACGTCATCTCGTCACTGG 398

DB 301 ACCCGATGGCGGACGCGCTTAACGGGATCATCCCGTTCGCTGACGTCATCTCGTCACTGG 360

QY 399 CAAACCAAGTCGGCGGCGACGCGTACAGGAACTGATTCGCGGAGGACCCCGGCC 458

DB 361 CAAACCAAGTCGGCGGCGACGCGTACAGGAACTGATTCGCGGAGGACCCCGGCC 420

QY 459 GAATTCATGTTGGATTTCGGGGCTTACACCGGAGATCAACTCCCGGAGATGACGCC 518

DB 421 GAATTCATGTTGGATTTCGGGGCTTACACCGGAGATCAACTCCCGGAGATGACGCC 480

QY 519 GCGCGGTTTCGGCTCGCTGTCGGCGCGCTCAGATGTCGGACAGCGTGGCGAGTAC 578

DB 481 GCGCGGTTTCGGCTCGCTGTCGGCGCGCTCAGATGTCGGACAGCGTGGCGAGTAC 540

QY 579 CTGTTTCGCGCGCTCGGCGTTTTCAGTTCGCTGTTCTGGGTTCTGACGGTGGGTCGTTGG 638

DB 541 CTGTTTCGCGCGCTCGGCGTTTTCAGTTCGCTGTTCTGGGTTCTGACGGTGGGTCGTTGG 600

QY 639 ATAGTTTCGTCGGCGGTCGATGTCGGCGCGCTCCGCGTATGTCGGTGGATGAC 698

DB 601 ATAGTTTCGTCGGCGGTCGATGTCGGCGCGCTCCGCGTATGTCGGTGGATGAC 660

QY 699 GTACCGCGGGCAGGCGCGCTGACCGCGCGCTCCAGGTCGCGGTTGTCGGCGGCGCTAC 758

DB 661 GTACCGCGGGCAGGCGCGCTGACCGCGCGCTCCAGGTCGCGGTTGTCGGCGGCGCTAC 720

QY 759 GAGACGGGTCGTCGGCTGACCGTCCCGCGCGCTGATCCCGGAGAACCGTGTGAACTG 818

DB 721 GAGACGGGTCGTCGGCTGACCGTCCCGCGCGCTGATCCCGGAGAACCGTGTGAACTG 780

QY 819 ATGATTCTGATAGCGACCAACCTCTTGGGGCAAAACACCCCGGATTCGGGTCAGAG 878

DB 781 ATGATTCTGATAGCGACCAACCTCTTGGGGCAAAACACCCCGGATTCGGGTCAGAG 840

879	QC	GAATACGGGAGATGGGCCCAAGACGCCCGCGGATGTTTGGCTAGCCCGCGCG	938
841	DB	GCAGATACGGCGAGATGGGCCCAAGACGCCCGCGGATGTTTGGCTAGCCCGCGCG	900
939	QC	ACGGCAACGGCGAGATGGGCCCAAGACGCCCGCGGATGTTTGGCTAGCCCGCGCG	998
901	DB	ACGGCAACGGCGAGATGGGCCCAAGACGCCCGCGGATGTTTGGCTAGCCCGCGCG	960
999	QC	GGTGGCTCTCGAGCAGCGCGCGCGGTGAGAGAGCCCTCCGACACCGCGCGCGGAC	1058
961	DB	GGTGGCTCTCGAGCAGCGCGCGCGGTGAGAGAGCCCTCCGACACCGCGCGGAC	1020
1059	QC	CAGTTGATGAACAATGTGCCCGAGCGCTGCAACAGCTGGCCCGACGCCACCGAGGAC	1118
1021	DB	CAGTTGATGAACAATGTGCCCGAGCGCTGCAACAGCTGGCCCGACGCCACCGAGGAC	1080
1119	QC	ACGCTCTTTCGAAGCTGGGTGGCTGGAAGACGGCTCTCGCGCATCGGTGCGCGATC	1178
1081	DB	ACGCTCTTTCGAAGCTGGGTGGCTGGAAGACGGCTCTCGCGCATCGGTGCGCGATC	1140
1179	QC	AGCAACATGTTGCGATGGCAACAAACCATGTGCGATGACCAACTCGGGTGTGCGATG	1238
1141	DB	AGCAACATGTTGCGATGGCAACAAACCATGTGCGATGACCAACTCGGGTGTGCGATG	1200
1239	QC	ACCAACACCTTGAGCTCGATGTTGAAGGCTTTGCTCCGCGCGCGCCCGACGCGGTG	1298
1201	DB	ACCAACACCTTGAGCTCGATGTTGAAGGCTTTGCTCCGCGCGCGCCCGACGCGGTG	1260
1299	QC	CAAAACGCGCGCAAAACGCGGTTCGCGCGATGAGCTCGCTGGGAGCTCGCTGGGTTCT	1358
1261	DB	CAAAACGCGCGCAAAACGCGGTTCGCGCGATGAGCTCGCTGGGAGCTCGCTGGGTTCT	1320
1359	QC	TCGGGTCTGGCGGTGGGTGGCGGCCAACTTCGGGTGCGGCGCGCTCGGTTCGTTG	1418
1321	DB	TCGGGTCTGGCGGTGGGTGGCGGCCAACTTCGGGTGCGGCGCGCTCGGTTCGTTG	1380
1419	QC	TCGGTCCGCGAGGCTGGCGCGCGCAACACGAGCTCACCCCGCGCGCGGCGCTG	1478
1381	DB	TCGGTCCGCGAGGCTGGCGCGCGCAACACGAGCTCACCCCGCGCGCGGCGCTG	1440
1479	QC	CCGCTGACGAGCTGACGCGCGCGGAAGAGGCGCGCGAGATGCTGGCGCGGCTG	1538
1441	DB	CCGCTGACGAGCTGACGCGCGCGGAAGAGGCGCGCGAGATGCTGGCGCGGCTG	1500
1539	QC	CCGCTGGGAGATGGCGCGCAGGCGCGGTGGTGGCTCAGTGGTGTGCTGCTGCTCG	1598
1501	DB	CCGCTGGGAGATGGCGCGCAGGCGCGGTGGTGGCTCAGTGGTGTGCTGCTGCTCG	1560
1599	QC	CCGCAACCTATGATGCGCATTTCCGGCAGCGCGCATATCGCCCGCGCGCTTG	1658
1561	DB	CCGCAACCTATGATGCGCATTTCCGGCAGCGCGCATATCGCCCGCGCGCTTG	1620
1659	QC	TCGCAGACCGGTTCCGCGATTTCCCGCGCTGCGCCCTCGACCGCTCCGCGATGGTCGC	1718
1621	DB	TCGCAGACCGGTTCCGCGATTTCCCGCGCTGCGCCCTCGACCGCTCCGCGATGGTCGC	1680
1719	QC	CAAGTGGGCGCACAGGTGGTCAACATCAACACCAACTGGGCTACAAACGCGGTGGC	1778
1681	DB	CAAGTGGGCGCACAGGTGGTCAACATCAACACCAACTGGGCTACAAACGCGGTGGC	1740
1779	QC	GCGGGACCGCATGCTCATGATCCCAACGGTGTGCTGTGATGACCAACACCGATGATC	1838
1741	DB	GCGGGACCGCATGCTCATGATCCCAACGGTGTGCTGTGATGACCAACACCGATGATC	1800
1839	QC	CGGGCGCACCGATCAATGCGTTAGCGTCCGCTCCGCGCAACCTACGCGCTGAT	1898
1801	DB	CGGGCGCACCGATCAATGCGTTAGCGTCCGCTCCGCGCAACCTACGCGCTGAT	1860
1899	QC	GTGTCGGGTATGACCGACCCAGATGTGCGGTGTGAGCTGCGGTCGCGGTGGC	1958
1861	DB	GTGTCGGGTATGACCGACCCAGATGTGCGGTGTGAGCTGCGGTCGCGGTGGC	1920
1959	QC	CTGCCGTCCGGCGGAGATCCGTTGGGCGGTTCGCGTGGAGCCCGCTGTCGCGATGGC	2018

Db	1921	CTGCGCGTGGCGCGGATCGTGGCGCGTGGCGGTGGTGAGCCCGTCTGTCGCGATGGCG	1980
Qy	2019	AACAGCGGTGGCGAGGGCGGAACGCCCGGTGGCGGTGGTGGCGAGGGTGGTGGCGTGGCG	2078
Db	1981	AACAGCGGTGGCGAGGGCGGAACGCCCGGTGGCGGTGGTGGCGAGGGTGGTGGCGTGGCG	2040
Qy	2079	CAAAACCGTGCAGGCGTTCGGATTTCGCTGACCGGTGCGGAAGACACATTGAACGGGTTGATC	2138
Db	2041	CAAAACCGTGCAGGCGTTCGGATTTCGCTGACCGGTGCGGAAGACACATTGAACGGGTTGATC	2100
Qy	2139	CAGTTCGATGCGCGCGATCAGACCCCGGTGATTCGGCGGGCCCGTCTCAACGGCCTAGGA	2198
Db	2101	CAGTTCGATGCGCGCGATCAGACCCCGGTGATTCGGCGGGCCCGTCTCAACGGCCTAGGA	2160
Qy	2199	CAGGTGCGTGGTATGAACACGGCCGGCGTCC	2228
Db	2161	CAGGTGCGTGGTATGAACACGGCCGGCGTCC	2190
RESULT	9		
AAD	47110		
ID	AAD47110	standard; DNA; 2808 BP.	
AC	XX	AAD47110;	
AC	XX		
DT	29-AUG-2003	(revised)	
DT	27-JAN-2003	(first entry)	
XX			
DE	Mycobacterium sp. MTB72F-Leishmania sp. MAPS	(aka r55f)-fusion DNA.	
XX			
KW	Vaccine; immunity; diagnostic agent; gene therapy; MTB72F; MAPS;		
KW	chimeric; gene; ds.		
XX			
OS	Mycobacterium sp.		
OS	Leishmania sp.		
OS	Chimeric.		
XX			
Key	Location/Qualifiers		
FF	4..2796		
CDS	/*tag= a		
FT	/product= "Mycobacterium sp. MTB72F-Leishmania sp. MAPS		
FT	(aka r55f) fusion protein"		
FT			
FT			

RESULT 9

AAD47110
ID AAD47110 standard: DNA: 2808 BP.

XX AAD47110;
AC

XX
DT 29-AUG-2003 (revised)

DT 27-JAN-2003 (LTSC ENERGY) XX

DE
XX
Mycobacterium sp. M1B/ZF-1

KW vaccine; immunity;
KW chimeric: gene: ds.

XX
05 Mycobacterium sp.

OS Leishmania sp.
OS Chimeric

XX	Key
FU	T

FT	CDS	FT
4		

EE E

2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

XX

XX

XX

XX

XX
XX

DR P-PSDB; AAE29731.

PT New recombinant r

PT against pathogen

XX
F1 CURECULOSIS:

XX
XX
XX

CC The inversion region
CC fusion polypeptide

CC *Mycobacterium* sp.
CC *Mycobacterium* sp.

CC poly(phenylene oxide) from TSA, LeIF, N

CC are used in mechanism

100

```

/*tag= a
/product= "Mycobacterium sp. MTB72F-Leishmania sp. MAPS
(aka r95f) fusion protein"

```

PN WO200272792-A2.

19-SEP-2002.

AA
PF 13-MAR-2002; 2002WO-US008223.

13-MAR-2001: 2001US-0275837P.

PA (CORI-) CORIXA CORP.

XX
PT
Skeiky Y. Brannon M. Guderian J:

XX
DP WPT: 2002-759844/82

DR P-PSDB; AAE29731.
YY

PT New recombinant nucleic acid molecule comprising a leishmanin gene
PT PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
PT tuberculosis.

Example 6; Page 128-129; 155pp; English.

xx The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, Leir, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are

CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC Mycobacterium sp. MTB72F-Leishmania sp. thiol specific antioxidant [TSA;
 CC MAPS (aka r95f)] fusion DNA. This sequence comprises Mycobacterium sp.
 CC MTB72F (a 72 kDa poly-protein fusion construct comprising Ral2-TbH9-Ra35)
 CC linked to the Leishmania TSA (MAPS). (Updated on 29-AUG-2003 to
 CC standardise OS field)

XX SQ Sequence 2808 BP; 466 A; 889 C; 982 G; 471 T; 0 U; 0 Other;

Query Match 95.6%; Score 2186.8; DB 6; Length 2808;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	39	CATATGATCAACCATCAACATCAACGCGCGCTCCGATTAATTCAGCTGTCCAGGGT	98
Db	1	CATATGATCAACCATCAACATCAACGCGCGCTCCGATTAATTCAGCTGTCCAGGGT	60
Qy	99	GCGCAGGATTCGCGATTCGCGATCGCGCAGCGATGCGATCGCGGCCAGATCCGATCG	158
Db	61	GCGCAGGATTCGCGATTCGCGATCGCGCAGCGATGCGATCGCGGCCAGATCCGATCG	120
Qy	159	GGTGGGGGTCAACCAACCGTTTCATATCGGGGTACCGGCTTCCGCTTGGGTGTGTC	218
Db	121	GGTGGGGGTCAACCAACCGTTTCATATCGGGGTACCGGCTTCCGCTTGGGTGTGTC	180
Qy	219	GACACACACGACGCGGACGATCGACGCGTGTGCGGAGCGCTCCGCGCGCAAGT	278
Db	181	GACACACGCGACGCGGACGATCGACGCGTGTGCGGAGCGCTCCGCGCGCAAGT	240
Qy	279	CTCGGCATCTCACCGCGGAGTATCACCGCGGTGACCGCGCTCCGATCAACTCGGC	338
Db	241	CTCGGCATCTCACCGCGGAGTATCACCGCGGTGACCGCGCTCCGATCAACTCGGC	300
Qy	339	ACCGGATGCGGAGCGGCTTAACGGGATCATCCCGGTGACGTCATCTCGGTGACCTGG	398
Db	301	ACCGGATGCGGAGCGGCTTAACGGGATCATCCCGGTGACGTCATCTCGGTGACCTGG	360
Qy	399	CAAAACAAATCGCGCGCACGCTACAGGAAAGTGCATTTGCCGAGGACCCCGGC	458
Db	361	CAAAACAAATCGCGCGCACGCTACAGGAAAGTGCATTTGCCGAGGACCCCGGC	420
Qy	459	GAATTCATGTGGATTTGCGGGGTTTACCAACGAGATCAACTCCCGGAGGATGTACGCC	518
Db	421	GAATTCATGTGGATTTGCGGGGTTTACCAACGAGATCAACTCCCGGAGGATGTACGCC	480
Qy	519	GGCCCGGTTTCGCGCTCGCTGGTGGCGCGGCTCAGATGTGGACAGCTGGCGATGAC	578
Db	481	GGCCCGGTTTCGCGCTCGCTGGTGGCGCGGCTCAGATGTGGACAGCTGGCGATGAC	540
Qy	579	CTGTGTTTCGCGCGGCTCGCGGTTTCAAGTCGGTGTCTGCGGTCTGACGGTGGGTCTGG	638
Db	541	CTGTGTTTCGCGCGGCTCGCGGTTTCAAGTCGGTGTCTGCGGTCTGACGGTGGGTCTGG	600
Qy	639	ATAGGTTCGTGGCGGTTCTGATGTGGCGCGGCTTCGCGGTATGTGGCGTGGATGAGC	698
Db	601	ATAGGTTCGTGGCGGTTCTGATGTGGCGCGGCTTCGCGGTATGTGGCGTGGATGAGC	660
Qy	699	GTCAACCGCGGGCAGGCGGAGCTGACCGCGCGGCTCCGCGGTTCGCGCGGCGCTAC	758
Db	661	GTCAACCGCGGGCAGGCGGAGCTGACCGCGCGGCTCCGCGGTTCGCGCGGCGCTAC	720
Qy	759	GAGACCGGTTATGGGTGAGCGGTTCGCGCGGCTTCGCGGTATGTGGCGTGGATGAGC	818
Db	721	GAGACCGGTTATGGGTGAGCGGTTCGCGCGGCTTCGCGGTATGTGGCGTGGATGAGC	780
Qy	819	ATGATTTCTGATACGACCAACCTTTGGGGCAAAACACCCCGCGATCGCGTCAACGAG	878
Db	781	ATGATTTCTGATACGACCAACCTTTGGGGCAAAACACCCCGCGATCGCGTCAACGAG	840

Qy	879	GCGGAATACGCGAGATGTGGGCCCAAGACGCGCGCGGATGTTTGGCTACGCGCGCG	938
Db	841	GCGGAATACGCGAGATGTGGGCCCAAGACGCGCGCGGATGTTTGGCTACGCGCGCG	900
Qy	939	AGGGCAGCGCGACGCGAGCTTGTGCGTTTCGAGGAGGCGCGGAGATGACACGCG	998
Db	901	AGGGCAGCGCGACGCGAGCTTGTGCGTTTCGAGGAGGCGCGGAGATGACACGCG	960
Qy	999	GGTGGGCTCTCTCAGCAGCGCGCGCTCGAGGAGGCTCCGACACCGCGCGCGGAGAAC	1058
Db	961	GGTGGGCTCTCTCAGCAGCGCGCGCTCGAGGAGGCTCCGACACCGCGCGCGGAGAAC	1020
Qy	1059	CAGTTGATGAACAATGTGCCCCAGCGGCTGCAACAGCTTGGGCCAGCCACGAGGGCACC	1118
Db	1021	CAGTTGATGAACAATGTGCCCCAGCGGCTGCAACAGCTTGGGCCAGCCACGAGGGCACC	1080
Qy	1119	ACGCGCTTCTTCCAAAGCTGGGTGCGCTTGGAGAGCGGTCTCGCGCATCGCTCGCGATC	1178
Db	1081	ACGCGCTTCTTCCAAAGCTGGGTGCGCTTGGAGAGCGGTCTCGCGCATCGCTCGCGATC	1140
Qy	1179	ACCAACATGTTGATGCGCAACCAACACATGTGATGACAACTCGGGTGTGTCGATG	1238
Db	1141	ACCAACATGTTGATGCGCAACCAACACATGTGATGACAACTCGGGTGTGTCGATG	1200
Qy	1239	ACCAACATGTTGATGCGCAACCAACAGGCTTGTGCGCGCGCGCGCGCGCGCGGTG	1298
Db	1201	ACCAACATGTTGATGCGCAACCAAGGCTTGTGCGCGCGCGCGCGCGCGCGGTG	1260
Qy	1299	CAAAACCGCGCGCAAAACCGGCTCGGGCGATGCTGTGGCGAGCTCGCTCGGTTCT	1358
Db	1261	CAAAACCGCGCGCAAAACCGGCTCGGGCGATGCTGTGGCGAGCTCGCTCGGTTCT	1320
Qy	1359	TGCGGCTCGGGGCTGGGCTCGCGCGCAACTTGGTGGCGGCGCTCGCTCGGTTCT	1418
Db	1321	TGCGGCTCGGGGCTGGGCTCGCGCGCAACTTGGTGGCGGCGCTCGCTCGGTTCT	1380
Qy	1419	TCGTTGCGCGAGGCTGGCGCGCGCGCAACAGGAGCTCACCCCGCGCGCGCGCGCTG	1478
Db	1381	TCGTTGCGCGAGGCTGGCGCGCGCGCAACAGGAGCTCACCCCGCGCGCGCGCTG	1440
Qy	1479	CGGTGACCAACCTGACAGCGCGCGGAAAGAGGCGCGGCGAGATGCTGGGCGGCTG	1538
Db	1441	CGGTGACCAACCTGACAGCGCGCGGAAAGAGGCGCGGCGAGATGCTGGGCGGCTG	1500
Qy	1539	CCGTTGGGCGAGATGGGCGCGCGCGCGTGTGGGCTCAGTGGTGTGCTGGGTTCCG	1598
Db	1501	CCGTTGGGCGAGATGGGCGCGCGCGCGTGTGGGCTCAGTGGTGTGCTGGGTTCCG	1560
Qy	1599	CCGCGACCTATGTATGTCGCGATCTTCGCGAGCGCGGAGATTCGCCCGCGCGCTTG	1658
Db	1561	CCGCGACCTATGTATGTCGCGATCTTCGCGAGCGCGGAGATTCGCCCGCGCGCTTG	1620
Qy	1659	TCGACAGACCGGTTTCGCGCGCTTCGCGCGCTCGCGCGCTCGCGCGCTCGCGCG	1718
Db	1621	TCGACAGACCGGTTTCGCGCGCTTCGCGCGCTCGCGCGCTCGCGCGCTCGCGCG	1680
Qy	1719	CAAGTGGGCGACAGGTGGTCAACATCAACCAAACTGGGCTACAAACCGCGTGGGC	1778
Db	1681	CAAGTGGGCGACAGGTGGTCAACATCAACCAAACTGGGCTACAAACCGCGTGGGC	1740
Qy	1779	GCGGGACCGGCTATCGATCGATCCCAACGCTGTGCTGCTGACCAACACCGTGTATC	1838
Db	1741	GCGGGACCGGCTATCGATCGATCCCAACGCTGTGCTGCTGACCAACACCGTGTATC	1800
Qy	1839	GCGGGCGCCACCGACATCAATGGGTTTCAGCGTTCGGGCTCCGGCCAAACCTACGCGCTCGAT	1898
Db	1801	GCGGGCGCCACCGACATCAATGGGTTTCAGCGTTCGGGCTCCGGCCAAACCTACGCGCTCGAT	1860
Qy	1899	GTGGTTCGGGTATGACCGCACCGGATGTGCGGTTGCTGAGCTGCGCGGTGCGGTGGC	1958
Db	1861	GTGGTTCGGGTATGACCGCACCGGATGTGCGGTTGCTGAGCTGCGCGGTGCGGTGGC	1920
Qy	1959	CTGCGCTCGCGCGGATCGGTGGCGCGCTCGCGGTTGGTGAGCGGCTCGGTGGGCG	2018

1059 CAGTTGATCAACAAATGTCGCCAGGCGCTGCAACAGCTGGCCACGACCCACGAGGGCACC 1118
1021 CAGTTGATCAACAAATGTCGCCAGGCGCTGCAACAGCTGGCCACGACCCACGAGGGCACC 1080
1119 AGCCCTTCTTCAAGCTGGGTGGCTGTGGAGACGGTCTCGCCCATCTGGTTCGCCGATC 1178
1081 AGCCCTTCTTCAAGCTGGGTGGCTGTGGAGACGGTCTCGCCCATCTGGTTCGCCGATC 1140
1179 AGCAACATGGTTCGATGGCCCAACAAACACATGTCGATGACCACTCGGGTGTGTCGATG 1238
1141 AGCAACATGGTTCGATGGCCCAACAAACACATGTCGATGACCACTCGGGTGTGTCGATG 1200
1239 ACCAACACCTTGGATCTCGATGTTGAAGGCTTTGCTCGGGGGGGCCGCGAGCCGGT 1298
1201 ACCAACACCTTGGATCTCGATGTTGAAGGCTTTGCTCGGGGGGGCCGCGAGCCGGT 1260
1299 CAAACCGGGCCCAAAACGGGGTTCGGGCGATGAGTCTGCTGGGACGCTCGCTGGTTCT 1358
1261 CAAACCGGGCCCAAAACGGGGTTCGGGCGATGAGTCTGCTGGGACGCTCGCTGGTTCT 1320
1359 TCGGGTCTGGGGCGGTGGGCTGGCCCAACTTGGTGGGGGGGCGCTCGGTTCGTTG 1418
1321 TCGGGTCTGGGGCGGTGGGCTGGCCCAACTTGGTGGGGGGGCGCTCGGTTCGTTG 1380
1419 TCGGTGCCCGAGGCTGGGCGCGCCGCAACAGGACAGTCAACCCGGGGGGCGGGCGCTG 1478
1381 TCGGTGCCCGAGGCTGGGCGCGCCGCAACAGGACAGTCAACCCGGGGGGCGGGCGCTG 1440
1479 CCGCTGACAGCTGACAGCGCGCGCGGAAAGAGGGCCCGGACAGTCTGGGGGGCGCTG 1538
1441 CCGCTGACAGCTGACAGCGCGCGCGGAAAGAGGGCCCGGACAGTCTGGGGGGCGCTG 1500
1539 CCGGTGGGGAGAGTGGGGCGCGCGGCTGGTGGGCTCACTGCTGCTGCTGGTTCGG 1598
1501 CCGGTGGGGAGAGTGGGGCGCGCGGCTGGTGGGCTCACTGCTGCTGCTGGTTCGG 1560
1599 CCGGACCTTATGATGCGGCATCTTCGGGACCGCGGATATCGCCCGCGGGCGCTG 1658
1561 CCGGACCTTATGATGCGGCATCTTCGGGACCGCGGATATCGCCCGCGGGCGCTG 1620
1659 TCGGAGGACCGGTTCCGCGCTTCCCGCGCTGCGCTGCGCTGCGCGATGCTGCC 1718
1621 TCGGAGGACCGGTTCCGCGCTTCCCGCGCTGCGCTGCGCTGCGCGATGCTGCC 1680
1719 CAGGTGGGGCACAAGTGGTCAACATCAACCAACTGGGCTACCAACACCGCGTGGG 1778
1681 CAGGTGGGGCACAAGTGGTCAACATCAACCAACTGGGCTACCAACACCGCGTGGG 1740
1779 GCGGGACCGGCATGTCATCGATCCCAACGGTGTGCTGCTGACCAACACCGATGATC 1838
1741 GCGGGACCGGCATGTCATCGATCCCAACGGTGTGCTGCTGACCAACACCGATGATC 1800
1839 GCGGGCCCAACCGACATCAATGGCTTCAGCGTGGGCTCCGGCCAAACCTACGGCGTGGAT 1898
1801 GCGGGCCCAACCGACATCAATGGCTTCAGCGTGGGCTCCGGCCAAACCTACGGCGTGGAT 1860
1899 GTGGTGGGTATGACCGACCGAGATGTCGGGTGCTGAGTGGCGGGTGGCGGTGGG 1958
1861 GTGGTGGGTATGACCGACCGAGATGTCGGGTGCTGAGTGGCGGGTGGCGGTGGG 1920
1959 CTGCGGTGGCGGCGATCGGTGGCGGCTCGCGGTTGGTGGCGGCTCGTGGCGATGGG 2018
1921 CTGCGGTGGCGGCGATCGGTGGCGGCTCGCGGTTGGTGGCGGCTCGTGGCGATGGG 1980
2019 AACGCGGTGGCGAGGCGGAAACCGCCGCTGGGTGCTGGAGAGGTGTGCGCTCGG 2078
1981 AACGCGGTGGCGAGGCGGAAACCGCCGCTGGGTGCTGGAGAGGTGTGCGCTCGG 2040
2079 CAAACCGTGGCGGCTCGGATTCGCTGACCGGTGGCGAGAGACATTTGAACGGGTGATC 2138
2041 CAAACCGTGGCGGCTCGGATTCGCTGACCGGTGGCGAGAGACATTTGAACGGGTGATC 2100
2139 CAGTTGATGCCCGGATCAGCCCGGTGATTCGGGCGGCGCGCTGCTCAACCGGCTAGGA 2198

Db 2101 CAGTTGATGCCCGGATCCAGCCCGGTGATTCGGGCGGCGCTGCTCAACGCTTAGGA 2160
Qy 2199 CAGGTGGTGGTATGAACACGCGCGGTCC 2228
Db 2161 CAGGTGGTGGTATGAACACGCGCGGTCC 2190
RESULT 11
ADA26363
ID ADA26363 standard; DNA; 3060 BP.
XX ADA26363;
AC
XX
DT 20-NOV-2003 (first entry)
XX
DE M. bovis MTB72F and 85b complex antigen (fusion MTB103F) DNA.
XX
KW ds; gene; fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
XX tuberculosis; tuberculostatic; gene therapy; vaccine; MTB72F; 85B.
OS Chimeric.
XX Mycobacterium bovis.
XX
PH Key Location/Qualifiers
FT CDS 4..3054
FT /*tag= a
FT /product= "MTB72F and 85b complex (fusion MTB103F)"
XX
PN MO2003070187-A2.
XX
PD 28-AUG-2003.
XX
PF 18-FEB-2003; 2003WO-US004903.
XX
PR 15-FEB-2002; 2002US-0357351P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Guderian J, Reed S;
XX
DR WPI; 2003-697554/66.
XX P-PSDB; ADA26370.
XX
PT New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX
PS Claim 84; Fig 11; 112pp; English.
XX
CC The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX
SQ Sequence 3060 BP; 503 A; 977 C; 1070 G; 510 T; 0 U; 0 Other;
Query Match 95.6%; Score 2186.8; DB 8; Length 3060;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 39 CATATGATCACCATCACCATCACAGCGCGCTCCGATAACTTCCAGTGTCCAGGGT 98
Db 1 CATATGATCACCATCACCATCACAGCGCGCTCCGATAACTTCCAGTGTCCAGGGT 60
Qy 99 GGGCAGGATTCGCCATTCGCGATCGGGCAGGCGATCGGGCCAGATCCCGATCG 158
Db 61 GGGCAGGATTCGCCATTCGCGATCGGGCAGGCGATCGGGCCAGATCCCGATCG 120

[illegible][illegible]

RESULT 12
ADA26362
ID ADA26362 standard; DNA; 3104 BP.
XX

AC ADA26362;
XX 20-NOV-2003 (first entry)
XX Mycobacterium MTB72F-hTCC#1 (fusion MTB102tm2F) protein, encoding DNA.
XX ds; gene; fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX Chimeric.
OS Mycobacterium sp.
XX Key Location/Qualifiers
FT CDS 4..3072
FT /*tag= a
FT /product= "MTB72F-hTCC#1 (fusion MTB102tm2F) protein"
XX WO2003070187-A2.
XX 28-AUG-2003.
XX 18-FEB-2003; 2003WO-US004903.
XX 15-FEB-2002; 2002US-0357351P.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Gudexian J, Reed S;
XX WPI; 2003-697554/66.
XX P-PSDB; ADA26369.
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX Claim 84; Fig 10; 112pp; English.
XX The invention relates to a novel nucleic acid encoding a fusion
XX polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
XX MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
XX A polypeptide of the invention has tuberculostatic activity. A
XX polynucleotide of the invention may have a use in gene therapy, and as a
XX vaccine. The methods and compositions of the present invention are useful
XX for diagnosing, preventing and/or treating tuberculosis infection. The
XX present sequence is used in the exemplification of the invention.
SQ Sequence 3104 BP; 510 A; 968 C; 1104 G; 522 T; 0 U; 0 Other;
Query Match 95.6%; Score 2186.8; DB 8; Length 3104;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2186; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 39 CATATGATCACCATCACCATCACCACGGCGCGTCCGATAAATCCAGCTGTCCAGGCT 98
Db 1 CATATGATCACCATCACCATCACCACGGCGCGTCCGATAAATCCAGCTGTCCAGGCT 60
QY 99 GGGCAGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 158
Db 61 GGGCAGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 120
QY 159 GGTGGGGGTTCACCCACCGTTTCATATCGGGCCCTACCGCCCTTCCTCGGCTTGGGTGTC 218
Db 121 GGTGGGGGTTCACCCACCGTTTCATATCGGGCCCTACCGCCCTTCCTCGGCTTGGGTGTC 180
QY 219 GACAAACCGGACCGGACGAGTCCACGCGTGTGTCGGGACGCGTCCGGCGGCAAGT 278
Db 181 GACAAACCGGACCGGACGAGTCCACGCGTGTGTCGGGACGCGTCCGGCGGCAAGT 240
QY 279 CTGGGATCTCCACCGGACGAGTTCACCGCGGTTCGACGGCGCTCCGATCAACTCGGCC 338
Db 241 CTGGGATCTCCACCGGACGAGTTCACCGCGGTTCGACGGCGCTCCGATCAACTCGGCC 300

QY 339 ACCGGATGGCGGACGCGCTTAACGGGCATCATCCCGGTGACGTATCTCGGTGACCTGG 398
Db 301 ACCGGATGGCGGACGCGCTTAACGGGCATCATCCCGGTGACGTATCTCGGTGACCTGG 360
QY 399 CAAACCAAGTCGGGCGGACGCGTACAGGAAACGTGACATTTGGCGGAGGACCCCGGCC 458
Db 361 CAAACCAAGTCGGGCGGACGCGTACAGGAAACGTGACATTTGGCGGAGGACCCCGGCC 420
QY 459 GAATTCATGGTGGATTTCGGGGCGTTACACGGGAGATCACTCGGAGGATGTAGCC 518
Db 421 GAATTCATGGTGGATTTCGGGGCGTTACACGGGAGATCACTCGGAGGATGTAGCC 480
QY 519 GCGCCGGGTTCGGCCCTCGCTCGTGGCCCGCGCTCAGATGTGGGACAGCGTGGCGAGTAC 578
Db 481 GCGCCGGGTTCGGCCCTCGCTCGTGGCCCGCGCTCAGATGTGGGACAGCGTGGCGAGTAC 540
QY 579 CTGTTTTTCGGCGCGTTCGGCGTTTCAGTTCGGTGTCTGGGTCTTGAACGTTGGGTCTGG 638
Db 541 CTGTTTTTCGGCGCGTTCGGCGTTTCAGTTCGGTGTCTGGGTCTTGAACGTTGGGTCTGG 600
QY 639 ATAGGTTTCGTTCGGCGGTCTGATGTGGCGCGCTCGCGCTATGTGGCGTGTAGTACG 698
Db 601 ATAGGTTTCGTTCGGCGGTCTGATGTGGCGCGGTCTGATGTGGCGTGTAGTACG 660
QY 699 GTACACCGGCGGACGCGGTGACGCGTCCCGCGCGGTTCGGGTTCGTGGCGGCGCTTAC 758
Db 661 GTACACCGGCGGACGCGGTGACGCGTCCCGCGCGGTTCGGGTTCGTGGCGGCGCTTAC 720
QY 759 GAGACGCGGTATGGCTGACGCGTCCCGCGCGGTTCGCGGAGAACCGTGTGAACTG 818
Db 721 GAGACGCGGTATGGCTGACGCGTCCCGCGCGGTTCGCGGAGAACCGTGTGAACTG 780
QY 819 ATGATTTCTGATAGGACCAACTCTTTGGGCAAAACACCCCGCGGATCGCGGTCAACGAG 878
Db 781 ATGATTTCTGATAGGACCAACTCTTTGGGCAAAACACCCCGCGGATCGCGGTCAACGAG 840
QY 879 GCCGATACGCGGATGTGGCCCAAGACCCCGCGGATGTTTGGCTAGCGCGCGCG 938
Db 841 GCCGATACGCGGATGTGGCCCAAGACCCCGCGGATGTTTGGCTAGCGCGCGCG 900
QY 939 ACCGCGACGCGGACGCGGTTCGCGTTCGAGGAGGCGCGGAGATGACACAGCGCG 998
Db 901 ACCGCGACGCGGACGCGGTTCGCGTTCGAGGAGGCGCGGAGATGACACAGCGCG 960
QY 999 GGTGGGTCTCTGAGACGCGCGGTTCGAGGAGGCTTCGACACCGCGCGCGGCAAC 1058
Db 961 GGTGGGTCTCTGAGACGCGCGGTTCGAGGAGGCTTCGACACCGCGCGCGGCAAC 1020
QY 1059 CAGTTGATGAACAATGTGCCCCAGCGCTGCAACAGCTGGCCAGCCACGACGGGAC 1118
Db 1021 CAGTTGATGAACAATGTGCCCCAGCGCTGCAACAGCTGGCCAGCCACGACGGGAC 1080
QY 1119 ACAGCTTCTTCAAGCTGGGTGGCTGTGGAAGACGCTTCGCGCGATTCGGTTCGCCGATC 1178
Db 1081 ACAGCTTCTTCAAGCTGGGTGGCTGTGGAAGACGCTTCGCGCGATTCGGTTCGCCGATC 1140
QY 1179 AGCAACATGGTTCGATGGCCCAACACACATGTCGATGACCACTCGGTGTGTCATG 1238
Db 1141 AGCAACATGGTTCGATGGCCCAACACACATGTCGATGACCACTCGGTGTGTCATG 1200
QY 1239 ACCAACACCTTTGAGCTCGATTTGAAGGCTTTTCTCCGGCGGCGCCCGCCAGGCGGTG 1298
Db 1201 ACCAACACCTTTGAGCTCGATTTGAAGGCTTTTCTCCGGCGGCGCCCGCCAGGCGGTG 1260
QY 1299 CAAACCGGCGCAAAACGGGTTCGGCGGATGAGCTCGTGGGACGCTCGCTGGGTCT 1358
Db 1261 CAAACCGGCGCAAAACGGGTTCGGCGGATGAGCTCGTGGGACGCTCGCTGGGTCT 1320
QY 1359 TCGGCTCTGGGCGGTGGGTTCGGCGCAACTTTGGGTTCGGGCGGCTCGCTCGGTTCGTTG 1418
Db 1321 TCGGCTCTGGGCGGTGGGTTCGGCGCAACTTTGGGTTCGGGCGGCTCGCTCGGTTCGTTG 1380

QY 519 GGGCCGGGTTCCGCTCCTGCTGTTGCGCCGCGCTCAGATGTGGACACGCTGGCGAGTAC 578
Db 481 GGGCCGGGTTCCGCTCCTGCTGTTGCGCCGCGCTCAGATGTGGACACGCTGGCGAGTAC 540
QY 579 CTGTTTTCGGCCGCTGCGGCTGTTTCACTCCGTTGCTGCGGCTCTGACGCTGGGGTCTG 638
Db 541 CTGTTTTCGGCCGCTGCGGCTGTTTCACTCCGTTGCTGCGGCTCTGACGCTGGGGTCTG 600
QY 639 ATAGGTTCTGTCGGCGGCTCTGATGCTGGCGCGCTCCTCGCGTATGTGGCGTGGATGAGC 698
Db 601 ATAGGTTCTGTCGGCGGCTCTGATGCTGGCGCGCTCCTCGCGTATGTGGCGTGGATGAGC 660
QY 699 GTACCGGGGAGGCGGAGCTGACCGCGCGCCAGCTCGGGTGTGCTGGCGGCGCTAC 758
Db 661 GTACCGGGGAGGCGGAGCTGACCGCGCGCCAGCTCGGGTGTGCTGGCGGCGCTAC 720
QY 759 GAGACGGCTATGGGCTGACGCTGCGCCCGCGCTGATCGCGAGAACCGTGTGAACCTG 818
Db 721 GAGACGGCTATGGGCTGACGCTGCGCCCGCGCTGATCGCGAGAACCGTGTGAACCTG 780
QY 819 ATGATTTCTGATAGCGACCACTCTTGGGGGCAAAACACCCCGCGGATCGCGTCAACGAG 878
Db 781 ATGATTTCTGATAGCGACCACTCTTGGGGGCAAAACACCCCGCGGATCGCGTCAACGAG 840
QY 879 GCCGAATACGGCGAGATGTGGGCCAAGACCGCGCGCGATGTTTGCTACGCGCGGCG 938
Db 841 GCCGAATACGGCGAGATGTGGGCCAAGACCGCGCGCGATGTTTGCTACGCGCGGCG 900
QY 939 ACGCGACGCGACGCGCACTTCTCCGTTTCAAGAGCGCGCGAGATGACACGCGG 998
Db 901 ACGCGACGCGACGCGCACTTCTCCGTTTCAAGAGCGCGCGAGATGACACGCGG 960
QY 999 GGTGGGCTCTCGAGCAGGCGCGCGCTCGAGAGGCTCTCGACACCGCGCGGCGAC 1058
Db 961 GGTGGGCTCTCGAGCAGGCGCGCGCTCGAGAGGCTCTCGACACCGCGCGGCGAAC 1020
QY 1059 CAGTTGATGAACAATGTGCCCGACGCGCTGCAACAGCTGGCGCCAGCCAGCGAGGCGAC 1118
Db 1021 CAGTTGATGAACAATGTGCCCGACGCGCTGCAACAGCTGGCGCCAGCCAGCGAGGCGAC 1080
QY 1119 ACGGCTTCTTCAAGCTGGTGGCTGCTGGAAGAGCTCTCGCGGATCGGTCGCGATC 1178
Db 1081 ACGGCTTCTTCAAGCTGGTGGCTGCTGGAAGAGCTCTCGCGGATCGGTCGCGATC 1140
QY 1179 AGCAACATGTTGCTGATGGCCCAACACACATGTCGATGACCAACTCGGCTGTGTCGATG 1238
Db 1141 AGCAACATGTTGCTGATGGCCCAACACACATGTCGATGACCAACTCGGCTGTGTCGATG 1200
QY 1239 ACCAACCTTGAAGCTGATGTTGAAGGCTTGTCTCCGCGCGCGCGCCAGCGCGT 1298
Db 1201 ACCAACCTTGAAGCTGATGTTGAAGGCTTGTCTCCGCGCGCGCGCCAGCGCGT 1260
QY 1299 CAACACCGCGCAAAACGCGCTCGGCGATGAGCTCGTGGCGAGCTCGCTGGGTTCT 1358
Db 1261 CAACACCGCGCAAAACGCGCTCGGCGATGAGCTCGTGGCGAGCTCGCTGGGTTCT 1320
QY 1359 TCGGCTCTGGCGCTGGGCTGGCGCGCAACTTGGGTCGCGCGCTCGTGGCTGCTG 1418
Db 1321 TCGGCTCTGGCGCTGGGCTGGCGCGCAACTTGGGTCGCGCGCTCGTGGCTGCTG 1380
QY 1419 TCGGTCGCGAGGCTGGCGCGGCGCAACACAGCTGACCCCGCGCGCGCGCGCTG 1478
Db 1381 TCGGTCGCGAGGCTGGCGCGGCGCAACACAGCTGACCCCGCGCGCGCGCGCTG 1440
QY 1479 CCGCTGACGAGCTGACACGCGCGCGCAAGAGGCGCGCGGAGATGCTGGCGCGGCTG 1538
Db 1441 CCGCTGACGAGCTGACACGCGCGCGCAAGAGGCGCGCGGAGATGCTGGCGCGGCTG 1500
QY 1539 CCGGTCGGGCGAGTGGGCGCGGCGCGGCTGGGCTGAGTGGTGTGCTGCTGCTG 1598
Db 1501 CCGGTCGGGCGAGTGGGCGCGGCGCGGCTGGGCTGAGTGGTGTGCTGCTGCTG 1560

QY 1599 CCGGACCCCTATGTGATGTCGCGCATCTCCGCGAGCGCGGATATCGCCCCCGGCTTG 1658
Db 1561 CCGGACCCCTATGTGATGTCGCGCATCTCCGCGAGCGCGGATATCGCCCCCGGCTTG 1620
QY 1659 TCGCAGGACCGGTTCCGCGACTTCCCGCGCTGCCCTTCGACCCGCTCGCGATGCTGCGC 1718
Db 1621 TCGCAGGACCGGTTCCGCGACTTCCCGCGCTGCCCTTCGACCCGCTCGCGATGCTGCGC 1680
QY 1719 CAAGTGGGGCCACAGTGGTCAACATCAACACCAAACTGGGCTACAACAACCGCGTGGC 1778
Db 1681 CAAGTGGGGCCACAGTGGTCAACATCAACACCAAACTGGGCTACAACAACCGCGTGGC 1740
QY 1779 GCGGGACCGGATCGATCGATCCCAACCGTGTGCTGACCAACAACCAACCGTGTATC 1838
Db 1741 GCGGGACCGGATCGATCGATCCCAACCGTGTGCTGACCAACAACCAACCGTGTATC 1800
QY 1839 GCGGGCCGACCGACATCAATCGCTTCAGCTCGCTCCGCCAACTACGGGCTCGAT 1898
Db 1801 GCGGGCCGACCGACATCAATCGCTTCAGCTCGCTCCGCCAACTACGGGCTCGAT 1860
QY 1839 GTGCTCGGATGACCGCACCCAGATGTCCGGTGTGCTGACGCTCGCGGTGCGGTTGGC 1958
Db 1861 GTGCTCGGATGACCGCACCCAGATGTCCGGTGTGCTGACGCTCGCGGTGCGGTTGGC 1920
QY 1959 CTGCTCGGCGGCGATCGGTGGCGGCTCGCGTGTGCTGAGCCCGTCTCGCGATGGC 2018
Db 1921 CTGCTCGGCGGCGATCGGTGGCGGCTCGCGTGTGCTGAGCCCGTCTCGCGATGGC 1980
QY 2019 AACCGCTGGCGAGGCGGAAACCGCCCGTGTGCTGACGCTGCGAGGTGCTGCGCTGGC 2078
Db 1981 AACCGCTGGCGAGGCGGAAACCGCCCGTGTGCTGACGCTGCGAGGTGCTGCGCTGGC 2040
QY 2079 CAACCCGTGACGCTCGGATTCGCTGACCGCTGCGAAGAGACATTGAACCGGTTGATC 2138
Db 2041 CAACCCGTGACGCTCGGATTCGCTGACCGCTGCGAAGAGACATTGAACCGGTTGATC 2100
QY 2139 CAGTTGATGCGCGATCCAGCCCGGATGTCGCGGCGGCGCGCTGCTCAACCGCTAGGA 2198
Db 2101 CAGTTGATGCGCGATCCAGCCCGGATGTCGCGGCGGCGCGCTGCTCAACCGCTAGGA 2160
QY 2199 CAGTGGTGGTATCAACACCGCGCGCTCC 2228
Db 2161 CAGTGGTGGTATCAACACCGCGCGCTCC 2190

RESULT 14

AAD47084

ID AAD47084 standard; DNA; 2190 BP.

XX AAD47084;

AC AAD47084;

XX 29-AUG-2003 (revised)

DT 27-JAN-2003 (first entry)

XX 29-AUG-2003 (revised)

XX 27-JAN-2003 (first entry)

XX Mycobacterium sp. MTB72FmutSA fusion protein encoding DNA.

XX Vaccine, immunity; diagnostic agent; gene therapy; TBH9; antigen;

KW Ra35MutSA; Ra12; MTB72MutSA; chimeric; gene; ds.

XX Mycobacterium sp.

OS Mycobacterium tuberculosis.

XX Chimeric.

OS Mycobacterium tuberculosis.

XX Chimeric.

XX Key

FH Location/Qualifiers

CDS 1..2190

FT /*tag= a

FT /product= "MTB72FmutSA fusion protein"

XX

XX WO200272792-A2.

XX 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US008223.

```
XX PR 13-MAR-2001; 2001US-0275837P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Brannon M, Guderian J;
XX DR WPI; 2002-759844/82.
XX DR P-PSDB; AAB29709.
XX PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
XX PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX PT tuberculosis.
XX PS Disclosure; Page 92-93; 155pp; English.
XX CC The invention relates to a recombinant nucleic acid molecule encoding a
XX CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX CC polynucleotide sequence encoding an antigen or an antigenic fragment from
XX CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX CC polypeptide or its fragment. The Leishmania polynucleotide is selected
XX CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
XX CC are used in methods for eliciting immune response in mammals. They are
XX CC useful as vaccines to elicit protective immunity against pathogenic
XX CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
XX CC polypeptides are used for enhancing the expression of polynucleotides, as
XX CC in vivo diagnostic agents and for raising antibodies in a non-human
XX CC animal. The invention is used in gene therapy. The present sequence is a
XX CC DNA encoding M7B72F fusion protein. This fusion protein comprises
XX CC Ra35Mata mutant protein and Ra12 protein from M. tuberculosis and TbH9
XX CC protein from Mycobacterium sp. (Updated on 29-AUG-2003 to standardise OS
XX CC field)
XX SQ Sequence 2190 BP; 340 A; 697 C; 797 G; 356 T; 0 U; 0 Other;

Query Match 95.5%; Score 2185.2; DB 6; Length 2190;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2187; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 42 ATGCATCACCATCACCATCACACGGCGGTCGCGATTAATTCAGTGTCCAGGGTGGG 101
DB 1 ATGCATCACCATCACCATCACACGGCGGTCGCGATTAATTCAGTGTCCAGGGTGGG 60
QY 102 CAGGATTCGCATTCGATCGGGCAGCGATCGCGATCGCGGCGAGATCCGATCGGCT 161
DB 61 CAGGATTCGCATTCGATCGGGCAGCGATCGCGATCGCGGCGAGATCCGATCGGCT 120
QY 162 GGGGGTCAACCGGTCATATCGGGCTTACCGGCTTCTCGGCTTGGTGTGTCGAC 221
DB 121 GGGGGTCAACCGGTCATATCGGGCTTACCGGCTTCTCGGCTTGGTGTGTCGAC 180
QY 222 AACACGGGCAACCGGCAACGAGTCCAAACGGGTGGTGGGAGCGCTCGGGGCAAGTCTC 281
DB 181 AACACGGGCAACCGGCAACGAGTCCAAACGGGTGGTGGGAGCGCTCGGGGCAAGTCTC 240
QY 282 GGCATCTCACCGGCAAGTATCACCGGTCAGCGGCTCCGATCACTCGGCAAC 341
DB 241 GGCATCTCACCGGCAAGTATCACCGGTCAGCGGCTCCGATCACTCGGCAAC 300
QY 342 GCGATGGCGGACGGCTTAAACGGGTCATCCCGGTGAAGTCACTCGGTGACCTGGCAA 401
DB 301 GCGATGGCGGACGGCTTAAACGGGTCATCCCGGTGAAGTCACTCGGTGACCTGGCAA 360
QY 402 ACCAAGTCGGGCGGACGGTACAGGGAACGTGACATTCGCGAGGAGCCCGCGGAA 461
DB 361 ACCAAGTCGGGCGGACGGTACAGGGAACGTGACATTCGCGAGGAGCCCGCGGAA 420
QY 462 TTCAATGTGGATTTCGGGGGGTTACCAACCGGAGATCAACTCCCGAGGATGTACGGCGG 521
DB 421 TTCAATGTGGATTTCGGGGGGTTACCAACCGGAGATCAACTCCCGAGGATGTACGGCGG 480
QY 522 CCGGGTTCGGCTCGCTGGTGGCGCGGCTCAGATGTGGACAGCGTGGCGAGTGACCTG 581
```

481 CCGGGTTCGGCTCGCTGGTGGCGCGGCTCAGATGTGGACAGCGTGGCGAGTGACCTG 540

582 TTTTCGGCGCGCTCGGCGGTTTCAAGTCGGTCTGGGGTCTGACGGTGGGGTGGTGGATA 641

541 TTTTCGGCGCGCTCGGCGGTTTCAAGTCGGTCTGGGGTCTGACGGTGGGGTGGTGGATA 600

642 GGTTCGTGGCGGGTCTGATGGTGGCGCGGCTCGCCGTATGTGGCGTGGATGAGCGTC 701

601 GGTTCGTGGCGGGTCTGATGGTGGCGCGGCTCGCCGTATGTGGCGTGGATGAGCGTC 660

702 ACCCGGGGAGCGCGAGCTGACCGCGCGCCAGCTCGGGTTCGTCGGCGGCTACGAG 761

661 ACCCGGGGAGCGCGAGCTGACCGCGCGCCAGCTCGGGTTCGTCGGCGGCTACGAG 720

762 ACGCGGTATGGGCTGACGGTTCGCCCGCGCGGTGATCCCGAGAACCTGTGTGAATGATG 821

721 ACGCGGTATGGGCTGACGGTTCGCCCGCGCGGTGATCCCGAGAACCTGTGTGAATGATG 780

822 ATTCGTATAGCGACCACTTTGGGCAAAACACCGCGCGATCGCGTCAACAGAGGCC 881

781 ATTCGTATAGCGACCACTTTGGGCAAAACACCGCGCGATCGCGTCAACAGAGGCC 840

882 GAATACGGCGAGATGTGGGCGCAAGACCGCGCGCGATGTTTGGCTACGCGCGCGGACG 941

841 GAATACGGCGAGATGTGGGCGCAAGACCGCGCGCGATGTTTGGCTACGCGCGCGGACG 900

942 GCGACGGCGAGCGGAGCTTCTCCGTTGAGAGAGCGCGCGAGATGACAGCGCGGCT 1001

901 GCGACGGCGAGCGGAGCTTCTCCGTTGAGAGAGCGCGCGAGATGACAGCGCGGCT 960

1002 GGGCTCTCGAGAGCGCGCGCGCTCGAGAGAGCTTCGACACCGCGCGCGGCGAACGAG 1061

961 GGGCTCTCGAGAGCGCGCGCGCTCGAGAGAGCTTCGACACCGCGCGCGGCGAACGAG 1020

1062 TTGATGAACAAATGTGCCCGAGCGCTGCAACAGCTGGCGCGCGAGCGAGGACACGAG 1121

1021 TTGATGAACAAATGTGCCCGAGCGCTGCAACAGCTGGCGCGCGAGCGAGGACACGAG 1080

1122 CTTTCTTCAAGCTGGGTGGCTGTGGAGAGCGTCTCGCGCATCGTCCCGCATCAGC 1181

1081 CTTTCTTCAAGCTGGGTGGCTGTGGAGAGCGTCTCGCGCATCGTCCCGCATCAGC 1140

1182 AACATGGTGTGATGGCGCAACCAACATGTCGATGACCAATCTGGGTGTGTCGATGAC 1241

1141 AACATGGTGTGATGGCGCAACCAACATGTCGATGACCAATCTGGGTGTGTCGATGAC 1200

1242 AACATGGTGTGATGGCGCAACCAACATGTCGATGACCAATCTGGGTGTGTCGATGAC 1301

1201 AACATGGTGTGATGGCGCAACCAACATGTCGATGACCAATCTGGGTGTGTCGATGAC 1260

1302 ACCCGGGCGCAAAACCGGGTCCGGCGATGAGCTCGTGGCGAGCTCGCTGGGTCTTTCG 1361

1261 ACCCGGGCGCAAAACCGGGTCCGGCGATGAGCTCGTGGCGAGCTCGCTGGGTCTTTCG 1320

1362 GGTCTGGCGGCTGGGGTGGCGCGCAACCTGGGTGGCGGCGCTCGGTGGGTGGTGGT 1421

1321 GGTCTGGCGGCTGGGGTGGCGCGCAACCTGGGTGGCGGCGCTCGGTGGGTGGTGGT 1380

1422 GTGCCGAGGCTGGCGCGCGCGCAACCGAGAGTCAACCGCGCGCGCGCGCGCGCGCG 1481

1381 GTGCCGAGGCTGGCGCGCGCGCAACCGAGAGTCAACCGCGCGCGCGCGCGCGCGCG 1440

1482 CTGACAGCTGACCGAGCGCGCGGAAAGAGGGCGCGCGCGAGATGCTGGCGGGGTGGCG 1541

1441 CTGACAGCTGACCGAGCGCGCGGAAAGAGGGCGCGCGCGAGATGCTGGCGGGGTGGCG 1500

1542 GTGGGGAGATGGGGCG 1601

1501 GTGGGGAGATGGGGCG 1560

1602 CGACCTGATGATCGCGCATCTCCGCGAGCGCGCGGATATCGCGCGCGCGCGCGCGCGCG 1661

Db 1561 CGACCTATGTGATCCGCAATTCTCCGGAGCCGCGATATCGCCCGCGCCCTTGTG 1620
 QY 1662 CAGGACCGGTTCCGCGACTTCCCGCGCTGCCCTCGACCCGTCGGGATGTCGCCCAA 1721
 Db 1621 CAGGACCGGTTCCGCGACTTCCCGCGCTGCCCTCGACCCGTCGGGATGTCGCCCAA 1680
 QY 1722 GTGGGCGCACAGGTGGTCAACATCAACCAAACTGGGCTACAAACACCGCGTGGGCGC 1781
 Db 1681 GTGGGCGCACAGGTGGTCAACATCAACCAAACTGGGCTACAAACACCGCGTGGGCGC 1740
 QY 1782 GGGACCGGCATCGTCAATCGATCCCAAGGTGCTGCTGACCAACACCAACGATCGG 1841
 Db 1741 GGGACCGGCATCGTCAATCGATCCCAAGGTGCTGCTGACCAACACCAACGATCGG 1800
 QY 1842 GGGCCACCGCATCAATCGCTTCAGCGTCCGGCTCCGGCCAACTACGGCGTGGATGTG 1901
 Db 1801 GGGCCACCGCATCAATCGCTTCAGCGTCCGGCTCCGGCCAACTACGGCGTGGATGTG 1860
 QY 1902 GTCGGGTATGACCGACCCAGATGTGCGGTGCTGACAGTCGCGGTGCGGTCGCGCTG 1961
 Db 1861 GTCGGGTATGACCGACCCAGATGTGCGGTGCTGACAGTCGCGGTGCGGTCGCGCTG 1920
 QY 1962 CGTCGGCGGCGATCGGTGCGCGGTGCGGTGCTGAGCCGCTCGTCGCGATGGGCAAC 2021
 Db 1921 CGTCGGCGGCGATCGGTGCGCGGTGCGGTGCTGAGCCGCTCGTCGCGATGGGCAAC 1980
 QY 2022 AGCGGTGGGCGAGGCGGAAACCCCGCTGCGGTGCTGCGAGGGTGTGCGGCTCGGCAA 2081
 Db 1981 AGCGGTGGGCGAGGCGGAAACCCCGCTGCGGTGCTGCGAGGGTGTGCGGCTCGGCAA 2040
 QY 2082 ACCGTGCAAGCGTGGATTCGCTGACCGGTGCGGAGAGACATTTGAACGGTTCATCCAG 2141
 Db 2041 ACCGTGCAAGCGTGGATTCGCTGACCGGTGCGGAGAGACATTTGAACGGTTCATCCAG 2100
 QY 2142 TTCGATGCGCGATCCAGCCCGGTGATTCGGCGGGCGCGCTGCTCAACGGCCCTAGGACAG 2201
 Db 2101 TTCGATGCGCGATCCAGCCCGGTGATTCGGCGGGCGCGCTGCTCAACGGCCCTAGGACAG 2160
 QY 2202 GTGGTCGGTATGAACACGGCGCGTCTAG 2231
 Db 2161 GTGGTCGGTATGAACACGGCGCGTCTAG 2190

RESULT 15

AAD28343

ID AAD28343 standard; DNA; 2190 BP.

AC AAD28343;

DT 22-APR-2002 (first entry)

XX Mycobacterium species MTB72FmutSA fusion protein encoding DNA.

KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB72FmutSA; Ra12-TbH9-Ra35MutSA;
 KW mutant; mutin; ds.

XX Mycobacterium sp.
 OS

PH Key Location/Qualifiers
 FT CDS 1..2190

FT /tag= a
 FT /product= "MTB72FmutSA fusion protein"

FT misc_feature 22..417

FT /note= "Ra12 DNA fragment"

FT misc_feature 424..1596

FT /tag= c

FT /note= "TBH9FL DNA fragment"

FT misc_feature 1603..2187

FT /tag= d

FT /note= "Ra35 DNA fragment"

FT mutation replace(2128, T)

FT /tag= e

XX WO200198460-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US019959.

XX 20-JUN-2000; 2000US-00597796.

XX 01-FEB-2001; 2001US-0265737P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Alderson M;

XX WPI; 2002-147798/19.

XX P-PSDB; AA17573.

XX Composition comprising MTB39 antigen and MTB32A antigen from
 MYcobacterium species, useful for eliciting immune response in a subject.
 Claim 81; Page 108-109; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
 MYcobacterium species antigens, nucleotides encoding them and
 compositions comprising such fusion proteins. The present invention
 particularly relates to nucleic acids encoding fusion proteins that
 include two or more individual M. tuberculosis antigens which increase
 the serological sensitivity of sera from individuals infected with
 tuberculosis and methods for their use in diagnosis, prevention and
 treatment of tuberculosis infection. Sequences of the invention are
 useful for eliciting an immune response in a mammal, e.g., human,
 immunised with BCG. They are useful in the diagnosis, treatment and
 prevention of Mycobacterium infection. The fusion proteins and the
 polynucleotides are useful as diagnostic tools in patients infected with
 Mycobacterium, in vitro and in vivo assays for detecting humoral
 antibodies or cell-mediated immunity against M. tuberculosis, for the
 diagnosis of an infection or monitoring of disease progression, as
 immunogens to generate or elicit a protective immune response in a
 patient and for raising anti-M. tuberculosis antibodies in a non-human
 animal. Sequences of the invention are also used as vaccines. MTB32A
 fusion proteins of the invention are useful as in vivo diagnostic agents
 for intradermal skin test. The present sequence is a DNA encoding
 CC Mycobacterium species MTB72FmutSA (Ra12-TbH9-Ra35MutSA) mutant fusion
 CC protein.

XX Sequence 2190 BP; 340 A; 697 C; 797 G; 356 T; 0 U; 0 Other;
 SQ

Query Match 95.5%; Score 2185.2; DB 6; Length 2190;
 Best Local Similarity 99.9%; Pred. No. 0;

Matches 2187; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 42 ATGCATCAACATCAACATCAACGCGCGTCCGATTAACCTCCAGCTGTCCACGGTGGG 101
 Db 1 ATGCATCAACATCAACATCAACGCGCGTCCGATTAACCTCCAGCTGTCCACGGTGGG 60

QY 102 CAGGATTTCGCATTCGCATCGGCGAGCGATCGGATCGGCGCCAGATCCGATCGGGT 161
 Db 61 CAGGATTTCGCATTCGCATCGGCGAGCGATCGGATCGGCGCCAGATCCGATCGGGT 120

QY 162 GGGGGTCAACCCAGTTCATATCGGCGCTACCGCTTCCTCGCTTGGGTGTTGTCGAC 221
 Db 121 GGGGGTCAACCCAGTTCATATCGGCGCTACCGCTTCCTCGCTTGGGTGTTGTCGAC 180

QY 222 AACACGCGACGCGGCGACGATCCAGCGTGTGCGAGCGCTCCGCGGCGAGTCTC 281
 Db 181 AACACGCGACGCGGCGACGATCCAGCGTGTGCGAGCGCTCCGCGGCGAGTCTC 240

QY 282 GGCATCTCCACCGCGAGCGTGAATCACCAGCGGTCCAGCGCGCTCCGATCAACTCGGCCACC 341
 Db 241 GGCATCTCCACCGCGAGCGTGAATCACCAGCGGTCCAGCGCGCTCCGATCAACTCGGCCACC 300

QY 342 GCGATGGGAGCGCGCTTAACGGGCGATCATCCCGGTGACGTCATCTCGGTGACCTGGCAA 401

Db 301 GCGATGGCGGACGCGCTTAAAGGCGCATCATCCGGTGACGTCTCTCGGTGACCTGGCAA 360
QY 402 ACCAAGTCGGCGGCGACGCGTACAGGAAACGTGAACATTGGCGAGGAGACCCCGGCGCGAA 461
Db 361 ACCAAGTCGGCGGCGACGCGTACAGGAAACGTGAACATTGGCGAGGAGACCCCGGCGCGAA 420
QY 462 TTCATGGTCGATTCGGGCGGTTACACCGGAGATCAACTCCGCGAGGATGTACGCCGCG 521
Db 421 TTCATGGTCGATTCGGGCGGTTACACCGGAGATCAACTCCGCGAGGATGTACGCCGCG 480
QY 522 CCGGGTTCGGCGCTCGTGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGAACCTG 581
Db 481 CCGGGTTCGGCGCTCGTGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGAACCTG 540
QY 582 TTTTCGGCGCGTGGCGGTTTCACTGGTGGTCTGGGGTCTGACCGTGGGGTCTGGGATA 641
Db 541 TTTTCGGCGCGTGGCGGTTTCACTGGTGGTCTGGGGTCTGACCGTGGGGTCTGGGATA 600
QY 642 GGTTCGTTCGGCGGTTCTGATGTGGCGCGGCTCGCCGTATGTGGCGTGGATGAGCGTC 701
Db 601 GGTTCGTTCGGCGGTTCTGATGTGGCGCGGCTCGCCGTATGTGGCGTGGATGAGCGTC 660
QY 702 ACCCGGGGAGCGGAGCTGACCGCGCGCGAGTCCGGGTTGCTCGCGCGGCTTACGAG 761
Db 661 ACCCGGGGAGCGGAGCTGACCGCGCGCGAGTCCGGGTTGCTCGCGCGGCTTACGAG 720
QY 762 ACCGCTATGGGTGACGCTGCGCGCGCGTATCGCGAGAACGCTGCTCAACTGATG 821
Db 721 ACCGCTATGGGTGACGCTGCGCGCGCGTATCGCGAGAACGCTGCTCAACTGATG 780
QY 822 ATTCTGATAGCGACCAACCTCTTTGGGCGAAACACCCCGCGCATCGCGTCAACGAGGCC 881
Db 781 ATTCTGATAGCGACCAACCTCTTTGGGCGAAACACCCCGCGCATCGCGTCAACGAGGCC 840
QY 882 GAATAGCGGAGATGTGGCGGCGGAGCGCGCGGATGTTGGCTACCGCGCGGCGAG 941
Db 841 GAATAGCGGAGATGTGGCGGCGGAGCGCGCGGATGTTGGCTACCGCGCGGCGAG 900
QY 942 GCGACGCGGAGCGGCGGCTGCTCGCGTTTCAGAGGCGCGCGAGATGACCGCGCGGT 1001
Db 901 GCGACGCGGAGCGGCGGCTGCTCGCGTTTCAGAGGCGCGCGAGATGACCGCGCGGT 960
QY 1002 GGGCTCTTCAGAGGCGCGCGGTCGAGGAGGCTTCGACACCGCGCGGCGGAGACCGAG 1061
Db 961 GGGCTCTTCAGAGGCGCGCGGTCGAGGAGGCTTCGACACCGCGCGGCGGAGACCGAG 1020
QY 1062 TTGATGAACATGTGCCCCAGGCGCTGCAACAGCTGGCCCCAGCCCCAGCGGSCACACG 1121
Db 1021 TTGATGAACATGTGCCCCAGGCGCTGCAACAGCTGGCCCCAGCCCCAGCGGSCACACG 1080
QY 1122 CCTTCTTCCAGCTGGGTGGCTGTGGAGACGCTTCGCCGATCGGTGCGCGATCAGC 1181
Db 1081 CCTTCTTCCAGCTGGGTGGCTGTGGAGACGCTTCGCCGATCGGTGCGCGATCAGC 1140
QY 1182 AACATGTGTGATGCGCAACACCAATGTCATGACCACTCGGGTGTGATGAC 1241
Db 1141 AACATGTGTGATGCGCAACACCAATGTCATGACCACTCGGGTGTGATGAC 1200
QY 1242 AACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCGCGCGGCGCGCGCGCGCGCGCG 1301
Db 1201 AACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCGCGCGGCGCGCGCGCGCGCG 1260
QY 1302 ACCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1361
Db 1261 ACCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
QY 1362 GGTCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1421
Db 1321 GGTCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
QY 1422 GTGCCGAGCGCTCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1481

Db 1381 GTGCCGAGCGCTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
QY 1482 CTGACGAGCGCTGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1541
Db 1441 CTGACGAGCGCTGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
QY 1542 GTGGGCGAGATGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1601
Db 1501 GTGGGCGAGATGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
QY 1602 CGACCTATGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1661
Db 1561 CGACCTATGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
QY 1662 CAGGACCGGTTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1721
Db 1621 CAGGACCGGTTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
QY 1722 GTGGGCGACAGGTGTGATCAACCAACCAACTGGGCTACACACGCGGCGGCGGCGGCGG 1781
Db 1681 GTGGGCGACAGGTGTGATCAACCAACCAACTGGGCTACACACGCGGCGGCGGCGGCGG 1740
QY 1782 GGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1841
Db 1741 GGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
QY 1842 GCGCGCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1901
Db 1801 GCGCGCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860
QY 1902 GTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1961
Db 1861 GTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
QY 1962 CCGTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2021
Db 1921 CCGTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1980
QY 2022 AGCGGTGGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2081
Db 1981 AGCGGTGGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040
QY 2082 ACCGTGACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2141
Db 2041 ACCGTGACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
QY 2142 TTCGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2201
Db 2101 TTCGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
QY 2202 GTGGTCCGCTATGAACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2231
Db 2161 GTGGTCCGCTATGAACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2190

Search completed: July 3, 2004, 17:52:08

Job time : 910 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2004, 17:23:11 ; Search time 169 Seconds
(without alignments)
7509.898 Million cell updates/sec

Title: US-09-597-796c-11

Perfect score: 2287
Sequence: 1 tctagaataattttgttta.....ggntgtaacaagccgaaa 2287

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2284	99.9	2287	4	US-09-223-040-1
2	2284	99.9	2287	4	US-09-287-849-1
3	1770.8	77.4	1797	4	US-09-287-849-25
4	1172.8	51.3	1801	4	US-09-287-849-21
5	1169	51.1	3058	3	US-08-818-112-106
6	1169	51.1	3058	4	US-08-818-111-101
7	1169	51.1	3058	4	US-09-056-556-106
8	1169	51.1	3058	4	US-09-072-596-101
9	1169	51.1	3058	4	US-09-072-967-106
10	1169	51.1	4403765	3	US-09-103-840A-2
11	1169	51.1	4411529	3	US-09-103-840A-1
12	947.4	41.4	4403765	3	US-09-103-840A-2
13	942.6	41.2	4411529	3	US-09-103-840A-1
14	941	41.1	3027	3	US-08-818-112-110
15	941	41.1	3027	4	US-08-818-111-105
16	941	41.1	3027	4	US-09-056-556-110
17	941	41.1	3027	4	US-09-072-596-105
18	941	41.1	3027	4	US-09-072-967-110
19	884	38.7	1725	3	US-08-818-112-108
20	884	38.7	1725	4	US-08-818-111-103
21	884	38.7	1725	4	US-09-056-556-108
22	884	38.7	1725	4	US-09-072-596-103
23	884	38.7	1725	4	US-09-072-967-108
24	718.8	31.4	851	3	US-08-818-112-33
25	718.8	31.4	851	4	US-08-818-111-33
26	718.8	31.4	851	4	US-09-056-556-33
27	718.8	31.4	851	4	US-09-072-596-33

US-09-223-040-1
; Sequence 1, Application US/09223040
; Patent No. 6545222
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-00901005
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein Rai2-TbH9-Ra35
; NAME/KEY: modified_base
; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
US-09-223-040-1

Query Match 99.9%; Score 2284; DB 4; Length 2287;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TCTAGAAATATTTGTTTACTTTAAGANGANGATATACATATGCATCCACCATCCATC 60
Db 1 TCTAGAAATATTTGTTTACTTTAAGANGANGATATACATATGCATCCACCATCCATC 60
Qy 61 ACACGGCCGGTCCGATAACTTCCAGGTGTCCAGGGTGGCAGGAGGATTCGCCATTCCGA 120
Db 61 ACACGGCCGGTCCGATAACTTCCAGGTGTCCAGGGTGGCAGGAGGATTCGCCATTCCGA 120
Qy 121 TCGGGGAGGAGTGGCGATCGCGGGCAGATCCGATCGGGTGGGGGTACCACCGTTTC 180

ALIGNMENTS

RESULT 1
US-09-223-040-1
; Sequence 1, Application US/09223040
; Patent No. 6545222
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-00901005
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein Rai2-TbH9-Ra35
; NAME/KEY: modified_base
; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
US-09-223-040-1

28 718.8 31.4 851 4 US-09-072-967-33 Sequence 33, Appl
29 588 25.7 1872 3 US-08-818-112-17 Sequence 17, Appl
30 588 25.7 1872 4 US-08-818-111-17 Sequence 17, Appl
31 588 25.7 1872 4 US-09-056-556-17 Sequence 17, Appl
32 588 25.7 1872 4 US-09-072-596-17 Sequence 17, Appl
33 588 25.7 1872 4 US-09-072-967-17 Sequence 17, Appl
34 426.6 18.7 702 4 US-09-287-849-27 Sequence 27, Appl
35 413 18.1 1012 4 US-09-643-597-351 Sequence 351, App
36 413 18.1 1012 4 US-09-606-421B-351 Sequence 351, App
37 398.2 17.4 447 3 US-08-818-112-4 Sequence 4, Appl
38 398.2 17.4 447 4 US-08-818-111-4 Sequence 4, Appl
39 398.2 17.4 447 4 US-09-056-556-4 Sequence 4, Appl
40 398.2 17.4 447 4 US-09-072-596-4 Sequence 4, Appl
41 398.2 17.4 447 4 US-09-072-967-4 Sequence 4, Appl
42 385.2 16.8 915 4 US-09-636-215-834 Sequence 834, App
43 385.2 16.8 915 4 US-09-685-166A-834 Sequence 834, App
44 384.4 16.8 675 4 US-09-636-215-822 Sequence 822, App
45 384.4 16.8 675 4 US-09-685-166A-822 Sequence 822, App

Db 121 TCGGCGAGCGGATGGCGATCGCGGCGCAGATCCGATCGGGTGGGGGTCAACCCACCGTTC 180
Qy 181 ATATCGGGCTACCGCTTCTCGGTTGGGTGGTGTGACAAACAGCGCAACGCGCAC 240
Db 181 ATATCGGGCTACCGCTTCTCGGTTGGGTGGTGTGACAAACAGCGCAACGCGCAC 240
Qy 241 GAGTCAACCGCTGTGGGAGCGCTCCGGCGGCAAGTCTCGGCATCTCCACCGCGGACG 300
Db 241 GAGTCAACCGCTGTGGGAGCGCTCCGGCGGCAAGTCTCGGCATCTCCACCGCGGACG 300
Qy 301 TGATCACCGCGTGAAGCGGCTCCGATCAACTCGGCCCAACCGCGATGCGGACGCGTTA 360
Db 301 TGATCACCGCGTGAAGCGGCTCCGATCAACTCGGCCCAACCGCGATGCGGACGCGTTA 360
Qy 361 ACGGCGATCATCCCGTGAGCGTCACTCGGTGACCTGCGGACCTGCAACCAAGTCCGGCGGACGC 420
Db 361 ACGGCGATCATCCCGTGAGCGTCACTCGGTGACCTGCGGACCTGCAACCAAGTCCGGCGGACGC 420
Qy 421 GTACAGGAACTGATGATGCGGAGGACCCCGCGCGAATTCATGTTGATTCGCGG 480
Db 421 GTACAGGAACTGATGATGCGGAGGACCCCGCGCGAATTCATGTTGATTCGCGG 480
Qy 481 CGTTACACCGGAGATCAACTCCGCGAGGATGACCGCGGCGCGGTTTGGCTCGCTGG 540
Db 481 CGTTACACCGGAGATCAACTCCGCGAGGATGACCGCGGCGCGGTTTGGCTCGCTGG 540
Qy 541 TGGCGCGGCTCAGATGTTGGGACAGCGTGGCGAGTGAACCTGTTTTCGCGCGCGTGGCGT 600
Db 541 TGGCGCGGCTCAGATGTTGGGACAGCGTGGCGAGTGAACCTGTTTTCGCGCGCGTGGCGT 600
Qy 601 TTGAGTGGTGGTCTGGGCTGACCGTGGGCTGAGTGGGTTCTGATGAGTTCGTCGCGGCTGA 660
Db 601 TTGAGTGGTGGTCTGGGCTGACCGTGGGCTGAGTGGGTTCTGATGAGTTCGTCGCGGCTGA 660
Qy 661 TGGTGGCGGCGCTCGCGGTATGTTGGCTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 720
Db 661 TGGTGGCGGCGCTCGCGGTATGTTGGCTGGATGAGCGTCAACCGCGGCGAGCGCGAGC 720
Qy 721 TGACCGCGCGCTCAGTGGGTTCTCGCGGCTTACGAGAGCGCGTATGGGCTGACGG 780
Db 721 TGACCGCGCGCTCAGTGGGTTCTCGCGGCTTACGAGAGCGCGTATGGGCTGACGG 780
Qy 781 TGCCCCCGCGGTGATCGCGAGAACGTTGAGTGAATCTGATAGCGACCAAC 840
Db 781 TGCCCCCGCGGTGATCGCGAGAACGTTGAGTGAATCTGATAGCGACCAAC 840
Qy 841 TCTTGGGGAACCAACCCCGCGATCGCGTCAACGAGGCGGAATACGGCGAGATGGG 900
Db 841 TCTTGGGGAACCAACCCCGCGATCGCGTCAACGAGGCGGAATACGGCGAGATGGG 900
Qy 901 CCCAAGAGCGCGCGATGTTGGCTACCGCGGCGGACGGCGAGCGCGGACG 960
Db 901 CCCAAGAGCGCGCGATGTTGGCTACCGCGGCGGACGGCGAGCGCGGACG 960
Qy 961 TGCTGCGCTTCGAGAGCGCGGAGATGACAGCGGCGGCTCTCGAGAGCGCG 1020
Db 961 TGCTGCGCTTCGAGAGCGCGGAGATGACAGCGGCGGCTCTCGAGAGCGCG 1020
Qy 1021 CCGCGTTCGAGAGCGCTCCGACACCGCGCGGCGAACCAGTTGATGAACAATGTCGCC 1080
Db 1021 CCGCGTTCGAGAGCGCTCCGACACCGCGCGGCGAACCAGTTGATGAACAATGTCGCC 1080
Qy 1081 AGGCGTTCGAACAGCTGGCGGCGGACCGCGAGGCGACCGCTCTTCCAGAGCTGGG 1140
Db 1081 AGGCGTTCGAACAGCTGGCGGCGGACCGCGAGGCGACCGCTCTTCCAGAGCTGGG 1140
Qy 1141 GCCTGTGGAAGAGCGTTCGCGCGATCGGTCGCGATGACGACATGTTGATGAGTGA 1200
Db 1141 GCCTGTGGAAGAGCGTTCGCGCGATCGGTCGCGATGACGACATGTTGATGAGTGA 1200
Qy 1201 ACAACCAATGTCGATGACCAACTCGGGTGTGTCGATGACCAACCTTGGCTGATGT 1260
Db 1201 ACAACCAATGTCGATGACCAACTCGGGTGTGTCGATGACCAACCTTGGCTGATGT 1260

Qy 1261 TGAAGGCTTTGTCTCGGCGGCGCGCCGACGCTGCAAAACCGCGGCGCAAAACGCGG 1320
Db 1261 TGAAGGCTTTGTCTCGGCGGCGCGCCGACGCTGCAAAACCGCGGCGCAAAACGCGG 1320
Qy 1321 TCCGCGCGATGAGTCTCGCTGGGCGAGCTCGCTGGGTCTTTTCGGGTCTGGGCGGTGG 1380
Db 1321 TCCGCGCGATGAGTCTCGCTGGGCGAGCTCGCTGGGTCTTTTCGGGTCTGGGCGGTGG 1380
Qy 1381 CCGCCAACTTGGGTCTGGGCGGCGCTCGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGG 1440
Db 1381 CCGCCAACTTGGGTCTGGGCGGCGCTCGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGG 1440
Qy 1441 CCGCCAACTTGGGTCTGGGCGGCGCTCGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGG 1500
Db 1441 CCGCCAACTTGGGTCTGGGCGGCGCTCGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGG 1500
Qy 1501 CCGCGAAAGAGAGGCGCGCGGCGAGATGCTGGGCGGCTGCGCGTGGGCGAGATGGGCGCA 1560
Db 1501 CCGCGAAAGAGAGGCGCGCGGCGAGATGCTGGGCGGCTGCGCGTGGGCGAGATGGGCGCA 1560
Qy 1561 GGGCGGCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGG 1620
Db 1561 GGGCGGCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGGGTCTGG 1620
Qy 1621 ATTCTCCGCGAGCGCGGATATCGCCCGCGCGGCTTGTGCGCGGCGAGCTTGTGCGCGG 1680
Db 1621 ATTCTCCGCGAGCGCGGATATCGCCCGCGCGGCTTGTGCGCGGCGAGCTTGTGCGCGG 1680
Qy 1681 TCCCGCGCTGCCCTCGACCCGCTCGCGATGTTGCCAAAGTGGGCGCACAGTGGTCA 1740
Db 1681 TCCCGCGCTGCCCTCGACCCGCTCGCGATGTTGCCAAAGTGGGCGCACAGTGGTCA 1740
Qy 1741 ACATCAACCAACCACTGGGCTACAAACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
Db 1741 ACATCAACCAACCACTGGGCTACAAACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
Qy 1801 ATCCCAACCGTGTCTGCTGACCAACCAACCAACCAACCAACCAACCAACCAACCA 1860
Db 1801 ATCCCAACCGTGTCTGCTGACCAACCAACCAACCAACCAACCAACCAACCAACCA 1860
Qy 1861 CGTTTCAGCGTGGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
Db 1861 CGTTTCAGCGTGGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
Qy 1921 AGGATGTCGCGGTGTGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1980
Db 1921 AGGATGTCGCGGTGTGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1980
Qy 1981 GCGGCGTTCGCGGTGTGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040
Db 1981 GCGGCGTTCGCGGTGTGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040
Qy 2041 CCGCGGTCGCGGTGTGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
Db 2041 CCGCGGTCGCGGTGTGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
Qy 2101 CGCTGACCGGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
Db 2101 CGCTGACCGGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
Qy 2161 CCGGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2220
Db 2161 CCGGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2220
Qy 2221 CCGGTCTTAGATATCCATCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2280
Db 2221 CCGGTCTTAGATATCCATCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2280
Qy 2281 CCGGAA 2287
Db 2281 CCGGAA 2287

RESULT 2
US-09-287-849-1
; Sequence 1, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein Ral2-TbH3-Ral5 (designated Mb332-Mb39
; NAME/KEY: modified_base
; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
US-09-287-849-1

Query Match 99.9%; Score 2284; DB 4; Length 2287;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TCTAGAAATAATTTTGTACTTTAAGAANGANATATACATATGCATCACCATC	60
Db	1	TCTAGAAATAATTTTGTACTTTAAGAANGANATATACATATGCATCACCATC	60
Qy	61	ACACGGCGCGTCCGATAAATTCAGCTGTCCAGGGTGGGCGAGGATTCGCCATTCCGA	120
Db	61	ACACGGCGCGTCCGATAAATTCAGCTGTCCAGGGTGGGCGAGGATTCGCCATTCCGA	120
Qy	121	TCGGGCGGCGATGGCGATTCGGGGCCAGATCCGATCGGTGGGGTCAACCCAGCTTC	180
Db	121	TCGGGCGGCGATGGCGATTCGGGGCCAGATCCGATCGGTGGGGTCAACCCAGCTTC	180
Qy	181	ATATCGGCGCTTACCGCTTCTCGGCTTGGGTGTGTGCGAACAAACGGCAACGGGCGAC	240
Db	181	ATATCGGCGCTTACCGCTTCTCGGCTTGGGTGTGTGCGAACAAACGGCAACGGGCGAC	240
Qy	241	GAGTCCAAACGGCTGTGGGAGCGCTCCGGGGCAAGTCTCGGATCTCCACCGGCGACG	300
Db	241	GAGTCCAAACGGCTGTGGGAGCGCTCCGGGGCAAGTCTCGGATCTCCACCGGCGACG	300

Qy	301	TGATCACCGCGGTGACGCGGCTCCGATCAATCGCCACCGCGATGGCGGACCGGCTTA	360
Db	301	TGATCACCGCGGTGACGCGGCTCCGATCAATCGCCACCGCGATGGCGGACCGGCTTA	360
Qy	361	ACGGGATCATCCGGTGACGTATCTCGTGACCTGGCAAAACCAAGTGGGGGGGACGC	420
Db	361	ACGGGATCATCCGGTGACGTATCTCGTGACCTGGCAAAACCAAGTGGGGGGGACGC	420
Qy	421	GTACAGGGAACGTGACATTCGGCGAGGACCCCGCGCCGAATTCATGGTGGATTTGGGG	480
Db	421	GTACAGGGAACGTGACATTCGGCGAGGACCCCGCGCCGAATTCATGGTGGATTTGGGG	480
Qy	481	CGTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGGCCCGGGTTCGGCTCGGTG	540
Db	481	CGTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGGCCCGGGTTCGGCTCGGTG	540
Qy	541	TGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGTTTTCGGCGCGCTCGGCGT	600
Db	541	TGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGTTTTCGGCGCGCTCGGCGT	600
Qy	601	TTCAAGTCGGTCTGGGGTCTGACCGTGGGGTCTGATAGTTCGTCTCGCGGGTCTGA	660
Db	601	TTCAAGTCGGTCTGGGGTCTGACCGTGGGGTCTGATAGTTCGTCTCGCGGGTCTGA	660
Qy	661	TGGTGGCGGCGCTCCCGCTATGTGCGGTGAGTGCACCGCGGGCGAGCCGAGC	720
Db	661	TGGTGGCGGCGCTCCCGCTATGTGCGGTGAGTGCACCGCGGGCGAGCCGAGC	720
Qy	721	TGACCGCGCCAGGTCCGGGTTCTCGCGGGCTACGAGACGGCGGTATGGGTGACGG	780
Db	721	TGACCGCGCCAGGTCCGGGTTCTCGCGGGCTACGAGACGGCGGTATGGGTGACGG	780
Qy	781	TGCCCCCGCGGTGATCCCGAGAACCGTGTCTGAACTGATGTTCTGATAGCAACCAACC	840
Db	781	TGCCCCCGCGGTGATCCCGAGAACCGTGTCTGAACTGATGTTCTGATAGCAACCAACC	840
Qy	841	TCCTGGGGCAAAACACACCCCGCGGATCGCGGTCAACAGAGGCCGAATACGGCAGATG	900
Db	841	TCCTGGGGCAAAACACACCCCGCGGATCGCGGTCAACAGAGGCCGAATACGGCAGATG	900
Qy	901	CCCAAGACCGCGCGCGATTTGGCTACGCGCGGGGACCGCGACGGCGACGGCGAGCT	960
Db	901	CCCAAGACCGCGCGCGATTTGGCTACGCGCGGGGACCGCGACGGCGACGGCGAGCT	960
Qy	961	TGCTCCGTTTCAGAGAGCGCGGAGATGACACGGGGTGGGCTCTTCAGCAGGCGCG	1020
Db	961	TGCTCCGTTTCAGAGAGCGCGGAGATGACACGGGGTGGGCTCTTCAGCAGGCGCG	1020
Qy	1021	CCGCGCTCAGAGAGCGCTCCGACACCGCGCGGGGAAACAGTTGATGAACAATGTCGCC	1080
Db	1021	CCGCGCTCAGAGAGCGCTCCGACACCGCGCGGGGAAACAGTTGATGAACAATGTCGCC	1080
Qy	1081	AGGCGCTGCAACAGCTGGCGGACCGCGACCGGGCACCCAGCTTCCTTCAAGCTGGGTG	1140
Db	1081	AGGCGCTGCAACAGCTGGCGGACCGCGACCGGGCACCCAGCTTCCTTCAAGCTGGGTG	1140
Qy	1141	GCTCTGGAAGACGTCTCGCGCATCGTCCCGCATCAGCAACATGGTGTGATGGGCA	1200
Db	1141	GCTCTGGAAGACGTCTCGCGCATCGTCCCGCATCAGCAACATGGTGTGATGGGCA	1200
Qy	1201	ACAACACATGTCGATGACCAACTCGGGTGTGTCATGACCAACACTTTAGCTCGATGT	1260
Db	1201	ACAACACATGTCGATGACCAACTCGGGTGTGTCATGACCAACACTTTAGCTCGATGT	1260
Qy	1261	TGAAGGGCTTTCCTCGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1320
Db	1261	TGAAGGGCTTTCCTCGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1320
Qy	1321	TCCGGCGGATGAGCTCGTGGGAGCTCGTGGGTTCCTTGGGTCTGGGCTGGGGGTGG	1380
Db	1321	TCCGGCGGATGAGCTCGTGGGAGCTCGTGGGTTCCTTGGGTCTGGGCTGGGGGTGG	1380

Db 564 GCTCTCAGCAGCGCGCGGCTCGAGGCGCTCCGACACCGCGCGGCGAACCAGTT 623
QY 1064 GATGAACAATGTGCGCCAGCGCGCTGCAACAGCTGCGCCAGCCAGCCAGCGCCAGCGC 1123
Db 624 GATGAACAATGTGCGCCAGCGCGCTGCAACAGCTGCGCCAGCCAGCCAGCGCCAGCGC 683
QY 1124 TTCTTCCAAAGTGGTGGCTGTGGAAGAGCGGTCTCGCGCATCGGTCGCGCATGAGCAA 1183
Db 684 TTCTTCCAAAGTGGTGGCTGTGGAAGAGCGGTCTCGCGCATCGGTCGCGCATGAGCAA 743
QY 1184 CATGTGTGATGCGCAACCAACCATGTGATGACCAACTCGGTTGTCGATGACCA 1243
Db 744 CATGTGTGATGCGCAACCAACCATGTGATGACCAACTCGGTTGTCGATGACCA 803
QY 1244 CACCTTGTAGCTCGATGTTGAAGGCTTGTCTCGCGCGCGCGCGCGCGCGCGCGCAAC 1303
Db 804 CACCTTGTAGCTCGATGTTGAAGGCTTGTCTCGCGCGCGCGCGCGCGCGCGCGCAAC 863
QY 1304 CGCGCGCAAAACCGGGTTCGGGCGATGAGTCTGCTGCGGAGTGTGCTGCGGTTTCGCG 1363
Db 864 CGCGCGCAAAACCGGGTTCGGGCGATGAGTCTGCTGCGGAGTGTGCTGCGGTTTCGCG 923
QY 1364 TCTGGCGGTTGGGTCGCGCAACTTGGGTCGGCGCGCGCTCGGTCGCTGCTGCTGCGGT 1423
Db 924 TCTGGCGGTTGGGTCGCGCAACTTGGGTCGGCGCGCGCTCGGTCGCTGCTGCTGCGGT 983
QY 1424 GCGCAGGCTTGGGTCGCGCAACCAAGGAGTCAACCGCGCGCGCGCGCGCGCGCGCT 1483
Db 984 GCGCAGGCTTGGGTCGCGCAACCAAGGAGTCAACCGCGCGCGCGCGCGCGCGCGCT 1043
QY 1484 GACGAGCTGACCGCGCGCGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1543
Db 1044 GACGAGCTGACCGCGCGCGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1103
QY 1544 GGGGAGATGGGCG 1603
Db 1104 GGGGAGATGGGCG 1163
QY 1604 ACCCTATGATGCGGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1663
Db 1164 ACCCTATGATGCGGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1223
QY 1664 GAGCGGTTCCCGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1723
Db 1224 GAGCGGTTCCCGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1283
QY 1724 GGGGCGAGGTGTCATCAACCAACCAAGTGGGTCGCGCGCGCGCGCGCGCGCGCG 1783
Db 1284 GGGGCGAGGTGTCATCAACCAACCAAGTGGGTCGCGCGCGCGCGCGCGCGCGCG 1343
QY 1784 GACGCGATGTCATCGATCCCAAGCGGTGCTGCTGACCAACCAACCAACCAACCGCG 1843
Db 1344 GACGCGATGTCATCGATCCCAAGCGGTGCTGCTGACCAACCAACCAACCAACCGCG 1403
QY 1844 GCGCACCGATCAATGCGTTGAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1903
Db 1404 GCGCACCGATCAATGCGTTGAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1463
QY 1904 GCGGTATGACCGCGACCGCGGATGTCGCGGTGCTGCGCGGTGCGCGGTGCGCGGTGCG 1963
Db 1464 GCGGTATGACCGCGACCGCGGATGTCGCGGTGCTGCGCGGTGCGCGGTGCGCGGTGCG 1523
QY 1964 GTCGCGCGGATCGGTGCGCGCGCGCGCGGTGCTGCGCGGTGCGCGGTGCGCGGTGCG 2023
Db 1524 GTCGCGCGGATCGGTGCGCGCGCGCGCGGTGCTGCGCGGTGCGCGGTGCGCGGTGCG 1583
QY 2024 GGTGGCGAGCGCGGAAACCGCGGTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2083
Db 1584 GGTGGCGAGCGCGGAAACCGCGGTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1643
QY 2084 GGTGAGCGGTCGATTCGCTGACCGGTGCGCGGAGAGACATGAAACGGGTTGATCCAGTT 2143
Db 1644 GGTGAGCGGTCGATTCGCTGACCGGTGCGCGGAGAGACATGAAACGGGTTGATCCAGTT 1703

QY 2144 CGATGCGCGGATCCAGCGCGCGGTGATTCGGCGCGCGCGCGCGCGCGCGCGCGCG 2203
Db 1704 CGATGCGCGGATCCAGCGCGCGGTGATTCGGCGCGCGCGCGCGCGCGCGCGCGCG 1763
QY 2204 GGTGCGGTATGAACACCGCGCGCGGTCTCTAGGATATC 2237
Db 1764 GGTGCGGTATGAACACCGCGCGCGGTCTCTAGGATATC 1797

RESULT 4

US-09-287-849-21
; Sequence 21, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1801
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein TbH9-DFV-MTI (designated Mtbf1f)
; NAME/KEY: CDS
; LOCATION: (1) .. (1800)
US-09-287-849-21

Query Match: 51.3%; Score 1172.8; DB 4; Length 1801;
Best Local Similarity 99.0%; Pred. No. 6e-225;
Matches 1180; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 464 CATGCTGATTTCCGGCGGTACCACCGGAGATCACTCCGCGAGATGTACCGCGGCC 523
Db 24 CATGCTGATTTCCGGCGGTACCACCGGAGATCACTCCGCGAGATGTACCGCGGCC 83
QY 524 GGGTTCCGCTCTGATGTCGCGCGCGCTCAGATGTGGGACAGCGTGGCGAGTACCTGTT 583
Db 84 GGGTTCCGCTCTGATGTCGCGCGCGCTCAGATGTGGGACAGCGTGGCGAGTACCTGTT 143
QY 584 TTCGCGCGCTCCGCTTTCAGTCCGCTGCTCGGGTCTGACGCTGGGTCTGATAGG 643
Db 144 TTCGCGCGCTCCGCTTTCAGTCCGCTGCTCGGGTCTGACGCTGGGTCTGATAGG 203
QY 644 TTCGTCGCGCGTCTGATGTCGCGCGCGCTCGCGCTATGTGCGGTGATGAGCTCAC 703
Db 204 TTCGTCGCGCGTCTGATGTCGCGCGCGCTCGCGCTATGTGCGGTGATGAGCTCAC 263
QY 704 CGCGGGGACGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 763
Db 264 CGCGGGGACGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 323

QY	764	GCGGTATGGGCTGA	CGGTGCCCGCCCGGTGATCCCGAGAACCGTGTGAACTGATGAT	823
DB	324	GGCGTATGGGCTGA	CGGTGCCCGCCCGGTGATCCCGAGAACCGTGTGAACTGATGAT	383
QY	824	TCTGATAGCGACCA	ACCTCTTCGGGCGAAACACCCCGCGATCGCGGTCAACGAGGCCGA	883
DB	384	TCTGATAGCGACCA	ACCTCTTCGGGCGAAACACCCCGCGATCGCGGTCAACGAGGCCGA	443
QY	884	ATACGGCGAGATGT	GGGCCCAAGACGCCCGCGGATGTTGGCTACTCGCCGGCGACGGC	943
DB	444	ATACGGCGAGATGT	GGGCCCAAGACGCCCGCGGATGTTGGCTACTCGCCGGCGACGGC	503
QY	944	GACGGCGAGCGAG	CGTGTGCTCCGTTTCGAGGAGCGCCGGAGATGACACAGCGCGGTGG	1003
DB	504	GACGGCGAGCGAG	CGTGTGCTCCGTTTCGAGGAGCGCCGGAGATGACACAGCGCGGTGG	563
QY	1004	GCTCTTCGAGCAGG	CCGCGCGGTTCGAGGAGGCTTCCGACACCGCGCGCGCAACCACTT	1063
DB	564	GCTCTTCGAGCAGG	CCGCGCGGTTCGAGGAGGCTTCCGACACCGCGCGCGCAACCACTT	623
QY	1064	GATGAACAATGTG	CCCCAGGCGCTGCAACAGTGTGCCCGAGCCCAACGAGGSCAACGACC	1123
DB	624	GATGAACAATGTG	CCCCAGGCGCTGCAACAGTGTGCCCGAGCCCAACGAGGSCAACGACC	683
QY	1124	TTCCTTCCAAGCT	GGGTGGCTTGTGAAGACGGTCTCGCCGCGATCGGTCCCGCATACACAA	1183
DB	684	TTCCTTCCAAGCT	GGGTGGCTTGTGAAGACGGTCTCGCCGCGATCGGTCCCGCATACACAA	743
QY	1184	CATGGTGTCTGAT	TGGCCAAACCAACATGTCTGATGACCAACTCGGGTGTGTCGATGACCAA	1243
DB	744	CATGGTGTCTGAT	TGGCCAAACCAACATGTCTGATGACCAACTCGGGTGTGTCGATGACCAA	803
QY	1244	CACCTTGAGCTCG	ATGTTGAAGGCTTTGCTCCGCGCGCGCCCGCCAGCGCGTGTGCAAC	1303
DB	804	CACCTTGAGCTCG	ATGTTGAAGGCTTTGCTCCGCGCGCGCCCGCCAGCGCGTGTGCAAC	863
QY	1304	CGCGCGCAAAAC	CGGGTCTCGGGCGATGAGTCTCGTGGGAGCTCGCTGGGTTCTTCGGG	1363
DB	864	CGCGCGCAAAAC	CGGGTCTCGGGCGATGAGTCTCGTGGGAGCTCGCTGGGTTCTTCGGG	923
QY	1364	TCTGGGCGGTGG	GGTGGCGCCCACTTCGGTTCGGGCGGCTCGGTGGTTCGTTCGGT	1423
DB	924	TCTGGGCGGTGG	GGTGGCGCCCACTTCGGTTCGGGCGGCTCGGTGGTTCGTTCGGT	983
QY	1424	GCOCAGGCTGG	CCGCGCCAACACAGGAGTCAACCCCGCGCGCGCGCGCTGCGCGT	1483
DB	984	GCOCAGGCTGG	CCGCGCCAACACAGGAGTCAACCCCGCGCGCGCGCGCTGCGCGT	1043
QY	1484	GACAGCGCTGAC	CGCCCGGAAAGAGGGCCCGGCGAGTGTGGCGGGCTGCGCGT	1543
DB	1044	GACAGCGCTGAC	CGCCCGGAAAGAGGGCCCGGCGAGTGTGGCGGGCTGCGCGT	1103
QY	1544	GGGGCAGATGG	GCGCCAGGCGCGGTGTGGGCTCAGTGGTGTCTCGTGTTCGCCCGCG	1603
DB	1104	GGGGCAGATGG	GCGCCAGGCGCGGTGTGGGCTCAGTGGTGTCTCGTGTTCGCCCGCG	1163
QY	1604	ACCCTATGATG	CGCGATCTCCGCGAGCCGCGGATATGCCCGCCCGCGCC	1655
DB	1164	ACCCTATGATG	CGCGATCTCCGCGAGCCGCGGATGATGCCCGCGGAC	1215

RESULT 5
US-08-818-112-106
; Sequence 106, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neco, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Wedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 153
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maxi, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 106:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3058 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-818-112-106

Query Match	51.1%	Score 1169	DB 3	Length 3058
Best Local Similarity	99.6%	Pred. No. 3,7e-224		
Matches 1172	Conservative 0	Mismatches 5	Indels 0	Gaps 0
QY	463	TCATGTGGGATTTCCGGGCGTTACCA	CCGAGATCAACTCCGCGAGGATCTAC	CGCGCC 522
Db	391	TAAATGGTGGGATTTCCGGGCGCTTACCA	CCGAGATCAACTCCGCGAGGATCTAC	CGCGCC 450
QY	523	CGGGTTCCGGCCTCGCTGGTGGCCCGCGCTCAGATGTGGGACACGCTGGCCAGTGACCTGT		582
Db	451	CGGGTTCCGGCCTCGCTGGTGGCCCGCGCTCAGATGTGGGACACGCTGGCCAGTGACCTGT		510
QY	583	TTTCGGCCCGGTCGGCGTTTCACTCGGTGTCTGCGGTCTGACGGTGGGTCTGGATG		642
Db	511	TTTCGGCCCGGTCGGCGTTTCACTCGGTGTCTGCGGTCTGACGGTGGGTCTGGATG		570
QY	643	GTTCGTCGGCGGTCTGATGTGGCGCGGCTTCGGCGTATGTGGCGTGATGAGCGTCA		702
Db	571	GTTCGTCGGCGGTCTGATGTGGCGCGGCTTCGGCGTATGTGGCGTGATGAGCGTCA		630
QY	703	CCGCGGGGAGGCCAGCTGACCGCCCGCCAGGTCGGGTGTGTCGGCGGCGCTACGAGA		762
Db	631	CCGCGGGGAGGCCAGCTGACCGCCCGCCAGGTCGGGTGTGTCGGCGGCGCTACGAGA		690
QY	763	CGGCGTATCGGCTGACGGTGTCGCCCGCGGTGATCCCGAGAACCGTGTGTAACCTGATGA		822
Db	691	CGGCGTATCGGCTGACGGTGTCGCCCGCGGTGATCCCGAGAACCGTGTGTAACCTGATGA		750
QY	823	TTCTGATACGACCAACCTCTTTGGGGCAAAA	CACCCCGCGGATCGCGGTCAACGAGCCG	882
Db	751	TTCTGATACGACCAACCTCTTTGGGGCAAAA	CACCCCGCGGATCGCGGTCAACGAGCCG	810
QY	883	AATACGGCGAGATGTGGGCCCAAGACGCCCGCGGATGTTTCGGCTACGCCGGCGGACCG		942
Db	811	AATACGGCGAGATGTGGGCCCAAGACGCCCGCGGATGTTTCGGCTACGCCGGCGGACCG		870
QY	943	CGACGGCGACCGGACGTTGTCGGCTGTCGAGGAGCGCCGAGATGACCA	CGCGGGTG 1002	
Db	871	CGACGGCGACCGGACGTTGTCGGCTGTCGAGGAGCGCCGAGATGACCA	CGCGGGTG 930	

1003 GGTCTCTCAGCAGCGCGCGGTTCGAGAGGCTCCGACACCGCGCGGGAACCACT 1062
Db GGTCTCTCAGCAGCGCGCGGTTCGAGAGGCTCCGACACCGCGCGGGAACCACT 990
1063 TGATGAACAATGTGCCAGCGGTTCGAAACAGTGGCCAGCCGACGAGGAGCACAACGC 1122
Db TGATGAACAATGTGCCAGCGGTTCGAAACAGTGGCCAGCCGACGAGGAGCACAACGC 1050
1123 CTTCTTCCAGCTGGGTGGCGGTTCGAAACAGTGGCCAGCGATGGTGGCGGATCAGCA 1182
Db CTTCTTCCAGCTGGGTGGCGGTTCGAAACAGTGGCCAGCGATGGTGGCGGATCAGCA 1110
1183 ACATGTGTGATGCGCAACCAACATGTGATGACCACTGGGTGGTGTGATGACCA 1242
Db ACATGTGTGATGCGCAACCAACATGTGATGACCACTGGGTGGTGTGATGACCA 1170
1243 ACACCTTGAGCTGATGTTGAAGGCTTTGCTCCGCGCGCGCGCGCGCGCGCGCA 1302
Db ACACCTTGAGCTGATGTTGAAGGCTTTGCTCCGCGCGCGCGCGCGCGCGCGCA 1230
1303 CCGCGCGCAAAACCGGCGTCCGCGCGATGAGCTCGCTGGGCGAGCTCGCTGGGTCTTCG 1362
Db CCGCGCGCAAAACCGGCGTCCGCGCGATGAGCTCGCTGGGCGAGCTCGCTGGGTCTTCG 1290
1363 GTCTGGCGGTGGGTGGCGCGCAACTGGGTGGCGCGCTCGCTGGGTGGGTGGGTGG 1422
Db GTCTGGCGGTGGGTGGCGCGCAACTGGGTGGCGCGCTCGCTGGGTGGGTGGGTGG 1350
1423 TGCGCGAGCGCTGGCGCGCGCAACAGCAGTCAACCGCGCGCGCGCGCGCGCGCG 1482
Db TGCGCGAGCGCTGGCGCGCGCAACAGCAGTCAACCGCGCGCGCGCGCGCGCGCG 1410
1483 TGACCGAGCTGACAGCGCGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1542
Db TGACCGAGCTGACAGCGCGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1470
1543 TGGGCGAGTGGCG 1602
Db TGGGCGAGTGGCG 1530
1603 GACCTATGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1639
Db GACCTATGTGATGCG 1567

RESULT 6
US-08-818-111-101
; Sequence 101, Application US/08818111
; Patent No. 633852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 101:
SEQUENCE CHARACTERISTICS:
LENGTH: 3058 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-101

Query March 51.1%; Score 1169; DB 4; Length 3058;
Best Local Similarity 99.6%; Pred. No. 3.7e-224;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 463 TCATGTGGATTTTCGGGCGGTACACCGAGATCACTCCGCGAGGATGTACCGCGCC 522
Db TAATGTGGATTTTCGGGCGGTACACCGAGATCAACTCCGCGAGGATGTACCGCGCC 450
QY 523 CGGTTTCGGCGCTTCGTCGCGCGCTCAGATGTGGAGCAGCGTGCAGTACACCTGT 582
Db CGGTTTCGGCGCTTCGTCGCGCGCTCAGATGTGGAGCAGCGTGCAGTACACCTGT 510
QY 583 TTTTCGGCGCGCTTCGTCGCGCTTCAGTTCGTTCTGGGCTCTGACGTTGGGCTGTGGATG 642
Db TTTTCGGCGCGCTTCGTCGCGCTTCAGTTCGTTCTGGGCTCTGACGTTGGGCTGTGGATG 570
QY 643 GTTCGTCGGCGCTTCGTCGCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTC 702
Db GTTCGTCGGCGCTTCGTCGCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTC 630
QY 703 CGCGCGCGCGCTTCGTCGCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTC 762
Db CGCGCGCGCGCTTCGTCGCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTC 690
QY 763 CGGCGTATGGGCTGACGTTCCGCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTCGCGCTTCG 822
Db CGGCGTATGGGCTGACGTTCCGCGCGCTTCGTCGCGCTTCGTCGCGCTTCGTCGCGCTTCG 750
QY 823 TTCTGATAGCGACCAACCTCTTGGGCGCAACACCGCGCGCTTCGTCGCGCTTCGTCGCGCTTC 882
Db TTCTGATAGCGACCAACCTCTTGGGCGCAACACCGCGCGCTTCGTCGCGCTTCGTCGCGCTTC 810
QY 883 AATACGCGAGATGTGGCGCGCTTCGAGAGGCGCTTCGACACCGCGCGCGCGCGCGCGCG 942
Db AATACGCGAGATGTGGCGCGCTTCGAGAGGCGCGCTTCGAGAGGCGCGCGCGCGCGCGCG 870
QY 943 CGACGCGAGCGCGAGCTTCGTCGCGCTTCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCG 1002
Db CGACGCGAGCGCGAGCTTCGTCGCGCTTCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCG 930
QY 1003 GGTCTCTCAGCAGCGCGCGCTTCGAGAGGCGCTTCGACACCGCGCGCGCGCGCGCGCG 1062
Db GGTCTCTCAGCAGCGCGCGCTTCGAGAGGCGCTTCGAGAGGCGCTTCGAGAGGCGCGCGCG 990
QY 1063 TGATGAACAATGTGCCAGCGGTTCGAAACAGTGGCCAGCCGACGAGGAGCACAACGC 1122
Db TGATGAACAATGTGCCAGCGGTTCGAAACAGTGGCCAGCCGACGAGGAGCACAACGC 1050
QY 1123 CTTCTTCCAGCTGGGTGGCGGTTCGAAACAGTGGCCAGCGATGGTGGCGGATCAGCA 1182
Db CTTCTTCCAGCTGGGTGGCGGTTCGAAACAGTGGCCAGCGATGGTGGCGGATCAGCA 1110
QY 1183 ACATGTGTGATGCGCAACCAACATGTGATGACCACTGGGTGGTGTGATGACCA 1242
Db ACATGTGTGATGCGCAACCAACATGTGATGACCACTGGGTGGTGTGATGACCA 1170

GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Hendrickson, Ronald C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 355
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/072,967
 FILING DATE: 05-MAY-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C9
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 106:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3058 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-09-072-967-106

Query Match	51.1%;	Score 1169;	DB 4;	Length 3058;
Best Local Similarity	99.6%;	Pred. No. 3.7e-224;		
Matches 1172;	Conservative	0;	Mismatches 5;	Indels 0;
Gaps	0;			

463 TCATGTTGGATTTCGGGGCGTTACACCGGAGATCAACTCGGAGATGATACCGCGCC 522
 391 TAATGTTGGATTTCGGGGCGTTACACCGGAGATCAACTCGGAGATGATACCGCGCC 450
 523 CGGGTTCCGGCTCGCTGGTGGCCCGCTCAGATGTGGGACAGCGTGGCGAGTGA 582
 451 CGGGTTCCGGCTCGCTGGTGGCCCGCTCAGATGTGGGACAGCGTGGCGAGTGA 510
 583 TTTCGGCGCGCTCGGGCTTCAGTCGCTGGTGTGGGTCTGACGGTGGGTCTGTGATG 642
 511 TTTCGGCGCGCTCGGGCTTCAGTCGCTGGTGTGGGTCTGACGGTGGGTCTGTGATG 570
 643 GTTCGTGGCGGCTCAGTGTGGCGCGCTCGCGGTATGTGGGTGGGTGATGAGGTCA 702
 571 GTTCGTGGCGGCTCAGTGTGGCGCGCTCGCGGTATGTGGGTGGGTGATGAGGTCA 630
 703 CCGCGGCGAGCGAGTCAAGCGCCCGCGAGTCCGGTGTGTGGCGCGCTTACGAGA 762
 631 CCGCGGCGAGCGAGTCAAGCGCCCGCGAGTCCGGTGTGTGGCGCGCTTACGAGA 690
 763 CGGGGTATGGCTACCGTCCCGCGGTGTGTCGGGAGAACCGTGTGTAATGATCA 822
 691 CGGGGTATGGCTACCGTCCCGCGGTGTGTCGGGAGAACCGTGTGTAATGATCA 750
 823 TTCTGATAGCAACCACTCTTGGGGCAAAACACCCCGCGATCGCGGTCAACGAGCGCG 882

RESULT 10

US-09-103-840A-2
 ; Sequence 2, Application US/09103840A
 ; Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 2

```
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      51.1%; Score 1169; DB 3; Length 4403765;
Best Local Similarity 99.6%; Pred. No. 1.2e-223;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 463 TCATGTGTGATTTTCGGGGGCTTACCAACCGAGATCAACTCCGAGAGATGTACGCGGCC 522
Db 1338813 TAATGTGTGATTTTCGGGGGCTTACCAACCGAGATCAACTCCGAGAGATGTACGCGGCC 1338872
QY 523 CGGGTTTCGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTGACCTGT 582
Db 1338873 CGGGTTTCGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTGACCTGT 1338932
QY 583 TTTCGCGCGGCTCGGCTTTCAGTCGGTGTCTGGGGTCTGACGCTGGGGTCTGGATAG 642
Db 1338933 TTTCGCGCGGCTCGGCTTTCAGTCGGTGTCTGGGGTCTGACGCTGGGGTCTGGATAG 1338992
QY 643 GTTCGTTCGGGGCTCTGATGGTGGCGCGCTCGCGGCTCTGGCGGTATGTGGGTGATGAGGTCA 702
Db 1338993 GTTCGTTCGGGGCTCTGATGGTGGCGCGCTCGCGGCTCTGGCGGTATGTGGGTGATGAGGTCA 1339052
QY 703 CCGCGGGGAGGCGGAGCTGACCGCGCGCCAGGTTCGGGGTTCGTCGGCGGCTTACGAGA 762
Db 1339053 CCGCGGGGAGGCGGAGCTGACCGCGCGCCAGGTTCGGGGTTCGTCGGCGGCTTACGAGA 1339112
QY 763 CGCGGTATGGCTGACGGTTCGCGCGCGCGCGGTGATCCGAGAACCGTGTGAACGTATGA 822
Db 1339113 CGCGGTATGGCTGACGGTTCGCGCGCGCGCGGTGATCCGAGAACCGTGTGAACGTATGA 1339172
QY 823 TTCTGATAGGACCAACTCTTGGGGCAAAAACACCCCGCGATCGCGGTCAACGAGCGCG 882
Db 1339173 TTCTGATAGGACCAACTCTTGGGGCAAAAACACCCCGCGATCGCGGTCAACGAGCGCG 1339232
QY 883 AATACGGCGAGATGTGGGCGCAAGACCGCGCGGATGTTTGGCTACGCGCGGCGAAGG 942
Db 1339233 AATACGGCGAGATGTGGGCGCAAGACCGCGCGGATGTTTGGCTACGCGCGGCGAAGG 1339292
QY 943 CGACGGCGAGGAGTGTGCTGCGCTTCGAGAGCGCGCGGAGATGACCGCGGGTG 1002
Db 1339293 CGACGGCGAGGAGTGTGCTGCGCTTCGAGAGCGCGCGGAGATGACCGCGGGTG 1339352
QY 1003 GGCTCTTCGAGCAGCGCGCGCGCTCGAGAGGCTTCCGACACCGCGCGGCGAACCAGT 1062
Db 1339353 GGCTCTTCGAGCAGCGCGCGCGCTCGAGAGGCTTCCGACACCGCGCGGCGAACCAGT 1339412
QY 1063 TGATGAACATGTGCGCGAGCGCTGCAACAGCTGCGCGAGCGCGAGGCGACCAAGC 1122
Db 1339413 TGATGAACATGTGCGCGAGCGCTGCAACAGCTGCGCGAGCGCGAGGCGACCAAGC 1339472
QY 1123 CTTCTTCCAGCTCGGTGGCTGTGAAGAAGCGTCTCGCGGATCGGTCCCGGATCAGCA 1182
Db 1339473 CTTCTTCCAGCTCGGTGGCTGTGAAGAAGCGTCTCGCGGATCGGTCCCGGATCAGCA 1339532
QY 1183 ACATGTGTGATGGCGCAACCAACATGTCGATGACCAACTCGGTGTGTGATGACCA 1242
Db 1339533 ACATGTGTGATGGCGCAACCAACATGTCGATGACCAACTCGGTGTGTGATGACCA 1339592
QY 1243 ACACCTTGAGCTGATGTTGAGGGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1302
Db 1339593 ACACCTTGAGCTGATGTTGAGGGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1339652
QY 1303 CCGCGCGCAAAAACGGGGTTCGGGCGATGAGCTCGCTGGGTCAGCTCGCTGGGTCTTCGG 1362
Db 1339653 CCGCGCGCAAAAACGGGGTTCGGGCGATGAGCTCGCTGGGTCAGCTCGCTGGGTCTTCGG 1339712
```

RESULT 11

```
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1
```

```
Query Match      51.1%; Score 1169; DB 3; Length 4411529;
Best Local Similarity 99.6%; Pred. No. 1.2e-223;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 463 TCATGTGTGATTTTCGGGGGCTTACCAACCGAGATCAACTCCGAGAGATGTACGCGGCC 522
Db 1339345 TAATGTGTGATTTTCGGGGGCTTACCAACCGAGATCAACTCCGAGAGATGTACGCGGCC 1339404
QY 523 CGGGTTTCGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTGACCTGT 582
Db 1339405 CGGGTTTCGGCTCGCTGGTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTGACCTGT 1339464
QY 583 TTTCGCGCGCTCGGCTTTCAGTCGGTGTCTGGGGTCTGACGCTGGGGTCTGGGATAG 642
Db 1339465 TTTCGCGCGCTCGGCTTTCAGTCGGTGTCTGGGGTCTGACGCTGGGGTCTGGGATAG 1339524
QY 643 GTTCGTTCGGCGGCTGATGTGGCGGGCGCTCGCGTATGTGGCGTGTGATGAGCGTCA 702
Db 1339525 GTTCGTTCGGCGGCTGATGTGGCGGGCGCTCGCGTATGTGGCGTGTGATGAGCGTCA 1339584
QY 703 CCGCGGGGAGCGCGAGCTGACCGCGCGAGGTTCGGGGTTCGCGCGCGCGCTTACGAGA 762
Db 1339585 CCGCGGGGAGCGCGAGCTGACCGCGCGAGGTTCGGGGTTCGCGCGCGCGCTTACGAGA 1339644
QY 763 CGCGGTATGGCTGACGGTTCGCGCGCGCGCGGTGATGCGCGAGAACCGTGTGAACTGATGA 822
Db 1339645 CGCGGTATGGCTGACGGTTCGCGCGCGCGCGGTGATGCGCGAGAACCGTGTGAACTGATGA 1339704
QY 823 TTCTGATAGGACCAACTCTTGGGGCAAAAACACCCCGCGGATCGCGTCAACGAGCGCG 882
```

```

Db 1339705 TTCTGATAGCACCACCTCTTGGGCAAAACACCCCGGCGATCGCGGTCAACGAGCGCG 1339764
QY 883 AATACGGCGAGATGGGCCCCAAGACCGCCCGGATGTTGGCTACGCGCGCGGCGACGG 942
Db 1339765 AATACGGCGAGATGGGCCCCAAGACCGCCCGGATGTTGGCTACGCGCGCGGCGACGG 1339824
QY 943 CGAGCGGACCGGACGCTTGTGCGGTTGAGAGAGGCGCCGGAGATGACACGCGCGGCGTG 1002
Db 1339825 CGAGCGGACCGGACGCTTGTGCGGTTGAGAGAGGCGCCGGAGATGACACGCGCGGCGTG 1339884
QY 1003 GGCTCTCGAGCAGCGCGCGCGTTCGAGAGGCGCTCGACACCGCGCGCGGCGAACGAGT 1062
Db 1339885 GGCTCTCGAGCAGCGCGCGCGTTCGAGAGGCGCTCGACACCGCGCGCGGCGAACGAGT 1339944
QY 1063 TGATGAACATGTGCCCCAGCGCTGCAACAGCTGGGCCAGCCACGACGCGGACCAACGCG 1122
Db 1339945 TGATGAACATGTGCCCCAGCGCTGCAACAGCTGGGCCAGCCACGACGCGGACCAACGCG 1340004
QY 1123 CTTCTTCAAGCTGGGTGGCTGTGGAAGAGCGTCTCGCGCATGCGTTCGCGCATGACCA 1182
Db 1340005 CTTCTTCAAGCTGGGTGGCTGTGGAAGAGCGTCTCGCGCATGCGTTCGCGCATGACCA 1340064
QY 1183 ACATGGTGTGATGGCAACCAACACATGTCGATGACCACTCGGCGTGTGTCGATGACCA 1242
Db 1340065 ACATGGTGTGATGGCAACCAACACATGTCGATGACCACTCGGCGTGTGTCGATGACCA 1340124
QY 1243 ACACCTTTGAGCTCGATGTTGAAGGCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCG 1302
Db 1340125 ACACCTTTGAGCTCGATGTTGAAGGCGCTTGTGCGCGCGCGCGCGCGCGCGCGCG 1340184
QY 1303 CCGCGGCGCAAAACGGGGTCCGGCGGATGAGCTGCTGGCAGCTGCTGGGTCTCTCGG 1362
Db 1340185 CCGCGGCGCAAAACGGGGTCCGGCGGATGAGCTGCTGGCAGCTGCTGGGTCTCTCGG 1340244
QY 1363 GTCTGGGCGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1422
Db 1340245 GTCTGGGCGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1340304
QY 1423 TGCGCGAGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1482
Db 1340305 TGCGCGAGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1340364
QY 1483 TGACGAGCTGACGAGCGCGCGGAGAGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1542
Db 1340365 TGACGAGCTGACGAGCGCGCGGAGAGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1340424
QY 1543 TGGGCGAGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1602
Db 1340425 TGGGCGAGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1340484
QY 1603 GACCTATGATGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1639
Db 1340485 GACCTATGATGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1340521

```

RESULT 12

```

US-09-103-840A-2/c
; Sequence 2, Application US/091039840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2

```

```

; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

Query Match 41.4%; Score 947.4; DB 3; Length 4403765;
Best Local Similarity 89.3%; Pred. No. 1.7e-179;
Matches 1062; Conservative 0; Mismatches 106; Indels 21; Gaps 3;

```

```

QY 466 TGGTGGATTTCCGGGCGTTACACCGGAGATCAACTCCGCGAGGATGTACCCGCGCGCGG 525
Db 1534826 TGGTGGATTTCCGGGCGTTACACCGGAGATCAACTCCGCGAGGATGTATCCGCGCGCGG 1534767
QY 526 GTTCCGGCTCTCGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTACCTGTTTTT 585
Db 1534766 GTTCCGGCTCTCGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTACCTGTTTTT 1534707
QY 586 CGGCGCGCTCGGCGGCTTTCACTCGGTGGTCTGGGCTCTGACGCGTGGGTCGTTGATAGTT 645
Db 1534706 CGGCGCGCTCGGCGGCTTTCACTCGGTGGTCTGGGCTCTGACGCGGATCGTTGATAGTT 1534647
QY 646 CGTCCGGCTCTGATGTGGCGCGGCTCGCGTATGTGGCGTGGATGAGCTGACCG 705
Db 1534646 CGTCCGGCTCTGATGTGGCGCGGCTCGCGTATGTGGCGTGGATGAGCTGACCG 1534587
QY 706 CGGCGCAGCGGAGCTGACCGCGCGGCTGAGTCCGGGTTGCTCGCGCGGCTTACGAGACGG 765
Db 1534586 CGGCGCAGCGGAGCTGACCGCGCGGCTGAGTCCGGGTTGCTCGCGCGGCTTACGAGACGG 1534527
QY 766 CGTATGGGCTGACGCTGCGCGCGGCTGATTCGCGGAGAACCGTCTGTAATGATGATTC 825
Db 1534526 CGTATGGGCTGACGCTGCGCGCGGCTGATTCGCGGAGAACCGTCTGTAATGATGATTC 1534467
QY 826 TGATAGCGACCAACTCTTGGGCGAAACACCCCGGCGATCGCGGTCAACGAGGCCGAAT 885
Db 1534466 TGATAGCGACCAACTCTTGGGCGAAACACCCCGGCGATCGCGGTCAACGAGGCCGAAT 1534407
QY 886 ACGGCGAGATGTGGGCGCAAGACCGCGCGCGATGTTGGTACGCGCGCGCGACGCGGA 945
Db 1534406 ACGGCGAGATGTGGGCGCAAGACCGCGCGCGATGTTGGTACGCGCGCGCGACGCGGA 1534347
QY 946 CGGCGACGCGGAGCTGCTGCGGTTTCGAGGAGGCGCGGAGATGACGAGCGCGGTCGGC 1005
Db 1534346 CGGCGACGCGGAGCTGCTGCGGTTTCGAGGAGGCGCGGAGATGACGAGCGCGGTCGGC 1534287
QY 1006 TCCTCGAGCAGCGCGCGCGCTCGAGGAGGCTTCGACACCGCGCGCGCGAACAGTTGA 1065
Db 1534286 TCCTTTCGAGCAGCGCGCTCGCGGTTTCGAGGAGGCGCGGAGATGACGAGCGCGGTC 1534227
QY 1066 TGACCAATGTGCCCCAGCGCTGCAACAGCTGGCGCGCGCGCGCGCGCGCGCGCGCTT 1125
Db 1534226 TGACCAATGTGCCCCAGCGCTGCAACAGCTGGCGCGCGCGCGCGCGCGCGCGCTT 1534167
QY 1126 CTTTCAAGCTGGGTGGCGCTGTGGAGAGCGGCTCTCGCGCGCATCGGTTCGCGCATGACCA 1185
Db 1534166 TCGACCAACTGAGTGAATCTTGGAAAGCGCATCTCGCGCGCATCTGTGCGCGCTCAGCAACA 1534107
QY 1186 TGGTGTGATGGCGCAACCAACCATGTCGATGACCACTCGGGTGTGTCGATGACCAACA 1245
Db 1534106 TGGTGTGATGTCACCAACCAACCGTGTGATGACCACTCGGGTGTGTCGATGACCAACA 1534047
QY 1246 CTTTCAGCTCGATGTTGAAGGCTTGTGCTCGGCGCGCGCGCGCGCGCGCGCGCGCG 1305
Db 1534046 CTTTCAGCTCGATGTTGAAGGCTTGTGCTCGGCGCGCGCGCGCGCGCGCGCGCGCG 1533990
QY 1306 CGGCGCAAAAACGGGGTCCGGGCGATGAGCTCGTGGGCGAGC-----TCGCTGG 1353
Db 1533989 CGGCGCAAAAACGGGGTCCAGGCGATGAGCTCGTGGGCGAGCAGCTGCGTGGTTCGCTGG 1533930

```

QY 1354 GTTCTCGGCTGTGGCGGTGGGGTGGCGCGCAACTTGGGTGGCGGCGCTCGGTGCTT 1413
Db 1533929 GTTCTTCGGGCTGTGGCGGTGGGGTGGCGCGCAACTTGGGTGGCGGCGCTCGGTGCTT 1533870
QY 1414 CGTTCGCTGTGGCGAGGCTGTGGCGCGCGCAACCCAGCAGTCAACCGCGCGCGCGG 1473
Db 1533869 CGTTCGCTGTGGCGAGGCTGTGGCGCGCGCAACCCAGCAGTCAACCGCGCGCGG 1533810
QY 1474 CGTTCGCTGTGGCGAGGCTGTGGCGCGCGCAACCCAGCAGTCAACCGCGCGCGG 1533
Db 1533809 CGTTCGCTGTGGCGAGGCTGTGGCGCGCGCAACCCAGCAGTCAACCGCGCGG 1533750
QY 1534 GGTTCGCGGTGTGGCGAGTGTGGCGCGCGCAACCCAGCAGTCAACCGCGCGG 1587
Db 1533749 GGTTCGCGGTGTGGCGAGTGTGGCGCGCGCAACCCAGCAGTCAACCGCGG 1533690
QY 1588 TGGTGTTCGCGCGAGTGTGGCGCGCGCAACCCAGCAGTCAACCGCGCGG 1536
Db 1533689 TGGTGTTCGCGCGAGTGTGGCGCGCGCAACCCAGCAGTCAACCGCGG 1533641

RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: PRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 41.2%; Score 942.6; DB 3; Length 4411529;
Best Local Similarity 89.1%; Pred. No. 1.5e-178;
Matches 1059; Conservative 0; Mismatches 109; Indels 21; Gaps 3;

QY 466 TGGTGGATTTTCGGGGGTTTACCAACCGGAGATCACTCCGGAGATGTACGGCGCGCG 525
Db 1533630 TGGTGGATTTTCGGGGGTTTACCAACCGGAGATCACTCCGGAGATGTACGGCGCGCG 1533571
QY 526 GTTTCGCGCTCGTGTGGCGCGCGCTCAGATGTGGGACAGCGTGGCGAGTCACTGTTT 585
Db 1533570 GTTTCGCGCTCGTGTGGCGCGCGCTCAGATGTGGGACAGCGTGGCGAGTCACTGTTT 1533511
QY 586 CGCGCGCGCTCGGTTTTCAGTTCGTGTGTGGGGTCTGAGCGTGGGGTCTGTGATAGTT 645
Db 1533510 CGCGCGCGCTCGGTTTTCAGTTCGTGTGTGGGGTCTGAGCGGATCGTGTGATAGTT 1533451
QY 646 CGTTCGCGGTCTGATGTGGCGCGCGCTTCGCGCTATGTGGCGTGGATGAGCGTCAACG 705
Db 1533450 CGTTCGCGGTCTGATGTGGCGCGCGCTTCGCGCTATGTGGCGTGGATGAGCGTCAACG 1533391
QY 706 CGGGGACGCGAGTCAACCGCGCGCGCGTTCGCGGTGTGGCGCGCGCTTACGAGCGG 765
Db 1533390 CGGGGACGCGAGTCAACCGCGCGCGCGTTCGCGGTGTGGCGCGCGCTTACGAGCGG 1533331
QY 766 CGTATGGGCTGAGCGTTCGCGCGCGCGTGTATCGCGGAGACCGTGTGACTGATGATTC 825
Db 1533330 CGTATGGGCTGAGCGTTCGCGCGCGCGTGTATCGCGGAGACCGTGTGACTGATGATTC 1533271
QY 826 TGATAGCGACCAACTCTTGGGGCAAAACACCCCGCGGATCGCGGTCAACGAGCGCGAAT 885

Db 1533270 TGATAGCGACCAACTCTTGGGGCAAAACACCCCGGCGATCGCGGTCAACGAGCGCGAAT 1533211
QY 886 AGCGGAGATGTGGGCCCAAGACCGCCCGCGATGTTTGGCTACGCGCGCGGACCGCGA 945
Db 1533210 AGCGGAGATGTGGGCCCAAGACCGCCCGCGATGTTTGGCTACGCGCGCGGACCGCGA 1533151
QY 946 CGGCGACGCGGACCGTTCGCGGTTCGAGGAGCGCGCGAGATGACACGAGCGCGGTGGC 1005
Db 1533150 CGGCGACGCGGAGCGTTCGCGGTTCGAGGAGCGCGCCACTGATCAACACCCCGCGCG 1533091
QY 1006 TCCTCGAGAGCGCGCGCGCGGTTCGAGAGCGCTTCGACACCGCGCGCGGCAACCGAGTTGA 1065
Db 1533090 TCCTTGAGAGCGCGCGCGGTTCGAGAGCGCTTCGACACCGCGCGCGGCAACCGAGTTGA 1533031
QY 1066 TGAACAATGTGCCCCAGCGCTGCAACAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1125
Db 1533030 TGAACAATGTGCCCCAGCGCTGCAACAACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1532971
QY 1126 CTTCCAGCTGGGTGGCGCTGTGGAGAGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 1185
Db 1532970 TCGACCAACTGAGTGAACCTCTGGAAGAGCGCTCTCGCGCGCGCGCGCGCGCGCGCG 1532911
QY 1186 TGGTGTGATGGCGCAACCAACATGTCGATGACCAACTCGGGTGTCTCGATGACCAACA 1245
Db 1532910 TCGTGTGATGCTCAACCAACCGCTGTCGATGACCAACTCGGGTGTCTCGATGACCAACA 1532851
QY 1246 CTTTGAGCTCGATGTTGAAGCGCTTTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1305
Db 1532850 CTTTGCACTCGATGTTGAAGCGCTTTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1532794
QY 1306 CGGCGCAAAACGGGCTCGGCGGATGAGCTCGCTGGCGAGC-----TCGCTGG 1353
Db 1532793 CGGCGCAAAACGGGCTCGGCGGATGAGCTCGCTGGCGAGCAGCTGGTTCGCTCGCTGG 1532734
QY 1354 GTTCTTCGGGTCTCGGCGGTGGCGGTGGCGCAACTTTGGGTTCGGCGCGCGCTTCGCTCGGT 1413
Db 1532733 GTTCTTCGGGTCTCGGCGGTGGCGGTGGCGCAACTTTGGGTTCGGCGCGCGCTTCGCTCGGT 1532674
QY 1414 GGTTCGCGTCCGCGAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1473
Db 1532673 GGTTCGCGTCCGCGAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1532614
QY 1474 CGTTCGCGTCCGCGAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1533
Db 1532613 CGTTCGCGTCCGCGAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1532554
QY 1534 GGTTCGCGTCCGCGAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1587
Db 1532553 GGTTCGCGTCCGCGAGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1532494
QY 1588 TGGTGTTCGCGCGCGCGCGCGCTATGTGATCGCGCGCGCGCGCGCGCGCGCGCGCG 1636
Db 1532493 TGGGATCGCGCGCGCGCGCGTACGTAATGCGCGCGCGCGCGCGCGCGCGCGCG 1532445

RESULT 14
US-08-818-112-110
; Sequence 110, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/818,112
 FILING DATE: 13-MAR-1997
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.41106
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 110:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3027 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-818-112-110

Query Match 41.1%; Score 941; DB 3; Length 3027;

Best Local Similarity 89.0%; Pred. No. 9.4e-179;

Matches 1058; Conservative 0; Mismatches 110; Indels 21; Gaps 3;

466 TGTGTGATTTCCGGGGCTTACACCGGAGATCAACTCCGGGAGATGACCGCGGCCCGG 525
 Db 1794 TGGTGGACTTCGGGGCGCTTACACCGGAGATCAACTCCGGGAGATGACCGCGGCCCGG 1853
 Qy 526 GTTCGGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCGTGGGAGTACTGTTTT 585
 Db 1854 GTTCGGCTCGCTGGTGGCGCGCGAGATGTGGGACAGCGTGGGAGTACTGTTTT 1913
 Qy 586 CGGCGCGCTCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 645
 Db 1914 CGGCGCGCTCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1973
 Qy 646 CGTGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 705
 Db 1974 CGTGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2033
 Qy 706 CGGCGGAGCGGAGCTGACCGCGCCCGGAGTCCGGGTGCTGGGGGGCTTACGAGACGG 765
 Db 2034 CGGCGGAGCGGAGCTGACCGCGCCCGGAGTCCGGGTGCTGGGGGGCTTACGAGACGG 2093
 Qy 766 CGTATGGCTGACGCTGCGCGCGCGCTGATCGCGAGACCGTGTGTAATGATGATTC 825
 Db 2094 CGTATGGCTGACGCTGCGCGCGCGCTGATCGCGAGAACCGTGTGTAATGATGATTC 2153
 Qy 826 TGATAGCGACCACTCTTGGGGGCAAAACACCGCGGATCGCGGTCAACGAGCGCGAAT 885
 Db 2154 TGATAGCGACCACTCTTGGGGGCAAAACACCGCGGATCGCGGTCAACGAGCGCGAAT 2213
 Qy 886 ACAGCGAGATGTGGGCGCAAGACCGCGCGGATGTTGGCTACGCGCGCGCGCGCGA 945
 Db 2214 ACAGCGAGATGTGGGCGCAAGACCGCGCGGATGTTGGCTACGCGCGCGCGCGCGA 2273
 Qy 946 CGGCGAGCGGACCTTCTCGCTCGCTCGAGAGCGCGCGAGATGACAGCGCGGGTGGGC 1005
 Db 2274 CGGCGAGCGGAGCTTCTCGCTCGCTCGAGAGCGCGCGAGATGACAGCGCGGGTGGGC 2333
 Qy 1006 TCCTTCGAGCGCGCGCGCTCGAGAGCGCTTCGACACCGCGCGCGCGCGCGCGCGCGA 1065
 Db 2334 TCCTTCGAGCGCGCGCTCGAGAGCGCTTCGACACCGCGCGCGCGCGCGCGCGCGCGA 2393

Qy 1066 TGAACAATGTGCCCGAGCGCTGCAACAGCTGGGCCAGCCAGCGGCGACCGCGCTT 1125
 Db 2394 TGAACAATGTGCCCGAGCGCTGCAACAGCTGGGCCAGCCAGCGGCGATCTGGCGGT 2453
 Qy 1126 CTTCCAGCTGGGTGGCTGTGGAGACGGTCTCGCGCATCGGTCCGCGATCAGCACCA 1185
 Db 2454 TCGACCAACTGAGTGAATCTTGAAAGCCATCTCGCGCATCTGTGCGCGCTCAGCAACA 2513
 Qy 1186 TGGTGTGATGCGCAACCAACCAACCATGTGATGACCAACTCGGGTGTGTGATGACCAACA 1245
 Db 2514 TGGTGTGATGCTCAACCAACCAACCATGTGATGACCAACTCGGGTGTGTGATGACCAACA 2573
 Qy 1246 CTTGAGCTCGATGTGAAAGGCTTGTCTCGGGGGGGGGCGCGAGCGCTGCAACCG 1305
 Db 2574 CTTGCACTCAATGTGAAAGGCTTGTCTCGGGGGGGGGCGCGAGCGCTGCAACCG 2630
 Qy 1306 CGGCGCAAAACGGGGTCCGGGCGATGAGCTCGCTGGGCGAGC-----TCGCTGG 1353
 Db 2631 CGGCGCAAAACGGGGTCCAGGGATGAGCTCGCTGGGCGAGCAGCTGGGTTCTGCTGG 2690
 Qy 1354 GTTCTTGGGTCTGGGCGGTGGGCGGCGCAACTTGGTCTGGGGGGCGCTCGGTGCGTT 1413
 Db 2691 GTTCTTGGGTCTGGGCGGTGGGCGGCGCAACTTGGTCTGGGGGGCGCTCGGTGCGTT 2750
 Qy 1414 GCTTGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1473
 Db 2751 GCTTGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2810
 Qy 1474 GCTTGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1533
 Db 2811 GCTTGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2870
 Qy 1534 GCTTGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1587
 Db 2871 GCTTGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2930
 Qy 1588 TGGTGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1636
 Db 2931 TGGGATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2979

RESULT 15

US-08-818-111-105

; Sequence 105, Application US/08818111

; Patent No. 6338852

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skelky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonia

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

; NUMBER OF SEQUENCES: 148

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/08/818,111

; APPLICATION NUMBER: 424

; FILING DATE: 13-MAR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

TUBERCULOSIS

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 3027 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-105

Query Match		41.1%; Score 941; DB 4; Length 3027;
Best Local Similarity		89.0%; Pred. No. 9,4e-179;
Matches 1058; Conservative		0; Mismatches 110; Indels 21; Gaps 3;
QY	466	TGTTGGATTTCCGGGCTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGCCCGG 525
DB	1794	TGTTGGATTTCCGGGCTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGCCCGG 1853
QY	526	GTTTCGGCTCGCTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGTTTT 585
DB	1854	GTTTCGGCTCGCTGGCGCGGCTGAGATGTGGGACAGCGTGGCGAGTGACCTGTTTT 1913
QY	586	CGCGCGCTCGCGCTTCACTCGTGGTCTGGGGTCTGACCGTGGGGTCTGGATAGTTT 645
DB	1914	CGCGCGCTCGCGCTTCACTCGTGGTCTGGGGTCTGACCGAGCGGATCGTGGATAGTTT 1973
QY	646	CGTCCGGGCTCTGATGTGGCGCGGCTCCCGCTGATGTGGCTGGATGACGTCACCG 705
DB	1974	CGTCCGGGCTCTGATGTGGCGCGGCTCCCGCTGATGTGGCTGGATGACGTCACCG 2033
QY	706	CGGGGAGCGCGAGCTGACCGCGCCCGAGTCCGGTTCCTGCGCGCGCTACAGACGG 765
DB	2034	CGGGGAGCGCGAGCTGACCGCGCCCGAGTCCGGTTCCTGCGCGCGCTACAGACGG 2093
QY	766	CGTATGGGCTGACGTGCCCCCGCGGTGATCGCGAGNACCGTCTGACTGATGATTC 825
DB	2094	CGTATGGGCTGACGTGCCCCCGCGGTGATCGCGAGNACCGTCTGACTGATGATTC 2153
QY	826	TGATAGGACCAACTCTTTGGGGGCAAAACACCCCGCGGATCGCGTCAACGAGGCGGAAT 885
DB	2154	TGATAGGACCAACTCTTTGGGGGCAAAACACCCCGCGGATCGCGTCAACGAGGCGGAAT 2213
QY	886	ACGGGAGATGTGGGCGCAAGACCGCGCGCGGATGTTGGCTACCGCGCGGCGAGCGGA 945
DB	2214	ACGGGAGATGTGGGCGCAAGACCGCGCGCGGATGTTGGCTACCGCGCGGCGAGCGGA 2273
QY	945	CGGCGAGCGCGAGCTTGTCTGCGTTCGAGGAGCGCGCGAGATGACCGCGCGGTGGCG 1005
DB	2274	CGGCGAGCGCGAGCTTGTCTGCGTTCGAGGAGCGCGCGAGATGACCGCGCGGTGGCG 2333
QY	1006	TCTCGAGAGCGCGCGCGGTGAGGAGCGCTTCGACACCGCGCGCGGCAACCAAGTTGA 1065
DB	2334	TCTCGAGAGCGCGCGCGGTGAGGAGCGCTTCGACACCGCGCGCGGCAACCAAGTTGA 2393
QY	1066	TGACCAATGTGCCAGGCGCTGCACAGCTGGCCCCGCGCCACGAGGCGCACCGCCTT 1125
DB	2394	TGACCAATGTGCCAGGCGCTGCACAGCTGGCCCCGCGCCACGAGGCGCACCGCCTT 2453
QY	1126	CTTCAAGTGGTGGCTGTGAGAGCGGTCTCGCGCGATCGGTCCGCGATCAGCAACA 1185
DB	2454	CTTCAAGTGGTGGCTGTGAGAGCGGTCTCGCGCGATCGGTCCGCGATCAGCAACA 2513
QY	1186	TGGTGTGATGGCGCAACCAACATGTCATGACCAACTCGGGTGTGTGATGACCAACA 1245
DB	2514	TGGTGTGATGGCGCAACCAACATGTCATGACCAACTCGGGTGTGTGATGACCAACA 2573
QY	1246	CCTTGAGCTCGATGTTGAAGGGCTTGTCTCGGCGCGCGCGCGCGCGCGCGCGCG 1305
DB	2574	CCTTGAGCTCGATGTTGAAGGGCTTGTCTCGGCGCGCGCGCGCGCGCGCGCGCG 2630

Search completed: July 3, 2004, 22:00:26
Job time : 207 secs

QY	1306	CGCGCAAAACCGGGTCCCGGGCGATGAGCTCGCTGGGCAAGC-----TCGCTGG 1353
DB	2631	CGCGCAAAACCGGGTCCAGGCGATGAGCTCGCTGGGCAAGCAGCTGGGTTCTCGCTGG 2690
QY	1354	GTTCTTCGGGTCTGGGCGGTGGGGTGGCGCAACTTGGTTCGGGGCGCTCGCTCGGTT 1413
DB	2691	GTTCTTCGGGTCTGGGCGGTGGGGTGGCGCAACTTGGTTCGGGGCGCTCGCTCGGTT 2750
QY	1414	CGTTTGTGGTGGCGCAGGCTTGGGCGCGCCCAACCAAGGAGTACCCCGGCGCGCGGG 1473
DB	2751	CGTTTGTGGTGGCGCAGGCTTGGGCGCGCCCAACCAAGGAGTACCCCGGCGCGCGGG 2810
QY	1474	CGTTCGGCTGACAGCCTGACCGAGCGCGGAGAGAGGCGCGGCGAGATGCTGGCG 1533
DB	2811	CGTTCGGCTGACAGCCTGACCGAGCGCGGAGAGAGGCGCGGCGAGATGCTGGCG 2870
QY	1534	GGCTTCGGGTGGGGCAGATGGGCGCCAGGCGCGG-----TGGTGGGCTCAGTGGTTC 1587
DB	2871	GGCTTCGGGTGGGGCAGATGGGCGCCAGGCGCGGTTTCGGCGCGGTTAGCAATGCGT 2930
QY	1588	TGGTGTTCGGCGCGGACCCCTATGATGCGCGATTTCTCCGGCAGCGGG 1636
DB	2931	TGGGATTCGGCGCGGCGGCTACGTAATGCCCGTGTGCCCGCGCGCGG 2979

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2004, 20:19:36 ; Search time 1007 Seconds
(without alignments)
10956.269 Million cell updates/sec

Title: US-09-597-796c-11

Perfect score: 2287

Sequence: 1 tctagaataatttggttta.....ggntgtaacaaagcccgaaa 2287

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3163042 seqs, 2412103800 residues

Total number of hits satisfying chosen parameters: 6326084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/2/pubpna/PCT05_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq2.*
- 14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2284	99.9	2287	9	US-09-287-849-1
2	2284	99.9	2287	12	US-09-886-349A-15
3	2284	99.9	2287	15	US-10-359-460-1
4	2284	99.9	2287	15	US-10-098-732A-15
5	2284	99.9	2287	16	US-10-359-459-1
6	2186.8	95.6	2451	16	US-10-369-983-8
7	2186.8	95.6	2487	16	US-10-369-983-7
8	2186.8	95.6	2637	16	US-10-369-983-6
9	2186.8	95.6	2808	15	US-10-098-732A-64
10	2186.8	95.6	2808	16	US-10-369-983-5
11	2186.8	95.6	3060	16	US-10-369-983-11
12	2186.8	95.6	3104	16	US-10-369-983-10
13	2186.8	95.6	3474	16	US-10-369-983-9
14	2185.2	95.5	2190	12	US-09-886-349A-17

15	2185.2	95.5	2190	15	US-10-098-732A-17	Sequence 17, Appl
16	1770.8	77.4	1797	9	US-09-287-849-25	Sequence 25, Appl
17	1770.8	77.4	1797	12	US-09-886-349A-19	Sequence 19, Appl
18	1770.8	77.4	1797	15	US-10-359-460-25	Sequence 25, Appl
19	1770.8	77.4	1797	15	US-10-098-732A-19	Sequence 19, Appl
20	1582	69.2	3030	16	US-10-369-983-3	Sequence 3, Appl
21	1572.6	68.9	2181	16	US-10-369-983-1	Sequence 1, Appl
22	1172.8	51.3	1801	9	US-09-287-849-21	Sequence 21, Appl
23	1172.8	51.3	1801	15	US-10-359-460-21	Sequence 21, Appl
24	1169	51.1	3058	12	US-09-886-349A-13	Sequence 13, Appl
25	1169	51.1	3058	15	US-10-193-002-101	Sequence 101, Appl
26	1169	51.1	3058	15	US-10-084-843-106	Sequence 106, Appl
27	1169	51.1	3058	15	US-10-098-732A-13	Sequence 13, Appl
28	1169	51.1	75216	15	US-10-080-170-646	Sequence 646, App
29	1169	51.1	75216	17	US-10-080-170-646	Sequence 646, App
30	941	41.1	3027	15	US-10-193-002-105	Sequence 105, App
31	941	41.1	3027	15	US-10-084-843-110	Sequence 110, App
32	921.8	40.3	1179	13	US-10-282-122A-25271	Sequence 26271, A
33	930.2	40.2	1182	13	US-10-282-122A-28708	Sequence 28708, A
34	884	38.7	1725	15	US-10-193-002-103	Sequence 103, App
35	884	38.7	1725	15	US-10-084-843-108	Sequence 108, App
36	718.8	31.4	851	12	US-09-886-349A-11	Sequence 11, Appl
37	718.8	31.4	851	15	US-10-193-002-33	Sequence 33, Appl
38	718.8	31.4	851	15	US-10-084-843-33	Sequence 33, Appl
39	718.8	31.4	851	15	US-10-098-732A-11	Sequence 11, Appl
40	591.2	25.9	1068	9	US-09-712-363-15	Sequence 15, Appl
41	589	25.8	1002	12	US-09-886-349A-5	Sequence 5, Appl
42	589	25.8	1002	15	US-10-098-732A-5	Sequence 5, Appl
43	588	25.7	1872	12	US-09-886-349A-1	Sequence 1, Appl
44	588	25.7	1872	15	US-10-193-002-17	Sequence 17, Appl
45	588	25.7	1872	15	US-10-084-843-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1

US-09-287-849-1
; Sequence 1, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein Ra12-TbH9-Ra35 (designated Mtb32-MCb39)
; OTHER INFORMATION: fusion)
; NAME/KEY: modified_base

```

; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
US-09-287-849-1

Query Match      99.9%; Score 2284; DB 9; Length 2287;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAGAAATAATTTTGTCTTAAAGAAAGANATATACATATCATCATCACCATCACCATC 60
DB 1 TCTAGAAATAATTTTGTCTTAAAGAAAGANATATACATATCATCATCACCATCACCATC 60

QY 61 ACACGGCCGCTCCGATTAATTCAGTGTCCAGGGTGGCCAGGATTCGCCATTCGGA 120
DB 61 ACACGGCCGCTCCGATTAATTCAGTGTCCAGGGTGGCCAGGATTCGCCATTCGGA 120

QY 121 TCGGSCAGCGATGCGGATCGGGCCAGATCCGATCGGGTGGGGTCCACCACCGTTC 180
DB 121 TCGGSCAGCGATGCGGATCGGGCCAGATCCGATCGGGTGGGGTCCACCACCGTTC 180

QY 181 ATATCGGCGCTACCGCTTCCTCGCTTGGTGTGTGCGAACAAACGGCAACGGCGCAC 240
DB 181 ATATCGGCGCTACCGCTTCCTCGCTTGGTGTGTGCGAACAAACGGCAACGGCGCAC 240

QY 241 GAGTCCAAACGGTGGTGGGAGCGCTCGGGCGGAAGTCTCGGATCTCCACGGCGAGC 300
DB 241 GAGTCCAAACGGTGGTGGGAGCGCTCGGGCGGAAGTCTCGGATCTCCACGGCGAGC 300

QY 301 TGATCACCGCGGTGCAAGCGGCTCCGATCAACTCGGCGCACCGGATGGCGGAGCGGCTTA 360
DB 301 TGATCACCGCGGTGCAAGCGGCTCCGATCAACTCGGCGCACCGGATGGCGGAGCGGCTTA 360

QY 361 ACGGGCAATATCCGGTGAAGTCACTCGGTGACCTGGGAAAACAAAGTGGGCGGACGC 420
DB 361 ACGGGCAATATCCGGTGAAGTCACTCGGTGACCTGGGAAAACAAAGTGGGCGGACGC 420

QY 421 GTACAGGGAACGTGACATTCGCGAGGACCCCGCGCGAATTCATGGTGGATTCGGGG 480
DB 421 GTACAGGGAACGTGACATTCGCGAGGACCCCGCGCGAATTCATGGTGGATTCGGGG 480

QY 481 CGTTACACCGGAGATCAACTCGCGAGGATGTACCGCGCGCGGCTTCGGGCTCGCTGG 540
DB 481 CGTTACACCGGAGATCAACTCGCGAGGATGTACCGCGCGCGGCTTCGGGCTCGCTGG 540

QY 541 TGCCCGGGGCTCAGATGCGGACAGTGGCGAGTGACCTGTTTCGGCGCGGCTCGGCGT 600
DB 541 TGCCCGGGGCTCAGATGCGGACAGTGGCGAGTGACCTGTTTCGGCGCGGCTCGGCGT 600

QY 601 TTCAGTCGGTGTGGGGTCTGACGCTGGGGTCTGGGATAGGTTCTGTCGGGGTCTGA 660
DB 601 TTCAGTCGGTGTGGGGTCTGACGCTGGGGTCTGGGATAGGTTCTGTCGGGGTCTGA 660

QY 661 TGTGGGGGGGCTCGCCGTATGTGGCTGGATGAGCGTACCGGGGGGCGAGCGGAGC 720
DB 661 TGTGGGGGGGCTCGCCGTATGTGGCTGGATGAGCGTACCGGGGGGCGAGCGGAGC 720

QY 721 TGACCGCGCGCCAGGTCGGGTTGCTGCGCGGCTACGAGACGGCGTATGGGCTGACGG 780
DB 721 TGACCGCGCGCCAGGTCGGGTTGCTGCGCGGCTACGAGACGGCGTATGGGCTGACGG 780

QY 781 TGCCCCCGCGGTGATCGCGGAAACCGTCTGAACTGATGATTCGTAGCGAACACC 840
DB 781 TGCCCCCGCGGTGATCGCGGAAACCGTCTGAACTGATGATTCGTAGCGAACACC 840

QY 841 TCTTGGGGCAAAACACCCCGGCGATCGGCTCAACGAGGCGCGAATACGCGGAGATGTGG 900

```

Db 1921 AGGATGTCGGGTGCTGCAAGCTGCGCGGTGCGGTGCGCTGCGTCCGCGCGGATCGGTG 1980
QY 1981 GCGCGTCCGGTGGTGGAGCCGCTGCTGCGATCGGGAACAGCGGTGGGAGCGCGAA 2040
Db 1981 GCGCGTCCGGTGGTGGAGCCGCTGCTGCGATCGGGAACAGCGGTGGGAGCGCGAA 2040
QY 2041 CCGCCGTCGGTGGTGGAGCGGTGCTGCGCTCGGCGCAACCGTGCAGGCGTGGATT 2100
Db 2041 CCGCCGTCGGTGGTGGAGCGGTGCTGCGCTCGGCGCAACCGTGCAGGCGTGGATT 2100
QY 2101 CGCTGACCGGTGCGGAGAGACATTGAACGGGTGATCCAGTTCGATCGCGCGATCCAGC 2160
Db 2101 CGCTGACCGGTGCGGAGAGACATTGAACGGGTGATCCAGTTCGATCGCGCGATCCAGC 2160
QY 2161 CCGGTGATTCGGCGCGCGCGCTGCTCAACGGGCTAGGACAGTGTTCGATGAACAGG 2220
Db 2161 CCGGTGATTCGGCGCGCGCGCTGCTCAACGGGCTAGGACAGTGTTCGATGAACAGG 2220
QY 2221 CCGGTGCTAGGATATCCATCACTGCGCGCGCTCGAGCAGATCCGCGTGTAAACAAG 2280
Db 2221 CCGGTGCTAGGATATCCATCACTGCGCGCGCTCGAGCAGATCCGCGTGTAAACAAG 2280
QY 2281 CCGGAAA 2287
Db 2281 CCGGAAA 2287

RESULT 2
US-09-886-349A-15
; Sequence 15, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Steven
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-00970US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; OTHER INFORMATION: MTB72F
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
; US-09-886-349A-15

Query Match 99.9%; Score 2284; DB 12; Length 2287;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAGAATAATTTGTTTACTTTTAAAGANGANATATACATATGCATACCATCACCATC 60
Db 1 TCTAGAATAATTTGTTTACTTTTAAAGANGANATATACATATGCATACCATCACCATC 60
QY 61 ACACGCGCGGTCCGATTAACCTTCCAGCTGTCCAGGGTGGGAGGGATTCCGCAATCCGA 120
Db 61 ACACGCGCGGTCCGATTAACCTTCCAGCTGTCCAGGGTGGGAGGGATTCCGCAATCCGA 120
QY 121 TCGGGAGGCGATGCGCATCGCGGCCAGATCCGATCGGGTGGGGTCCACCAACCGTTC 180
Db 121 TCGGGAGGCGATGCGCATCGCGGCCAGATCCGATCGGGTGGGGTCCACCAACCGTTC 180
QY 181 ATATCGGGCTTACCGCTTCTCGGCTTGGGTGTTCGACAAACGCGCAACGCGCGAC 240
Db 181 ATATCGGGCTTACCGCTTCTCGGCTTGGGTGTTCGACAAACGCGCAACGCGCGAC 240
QY 241 GAGTCCAAACGGGTGGTGGAGCGCTCCGGGCGAAGTCTCGGCATCTCCACCGGCGACG 300
Db 241 GAGTCCAAACGGGTGGTGGAGCGCTCCGGGCGAAGTCTCGGCATCTCCACCGGCGACG 300
QY 301 TGATCACCGCGGTGACGCGCTCCGATCAACTCGGCACCGCATGGCGAGCGCGTTA 360
Db 301 TGATCACCGCGGTGACGCGCTCCGATCAACTCGGCACCGCATGGCGAGCGCGTTA 360
QY 361 ACGGGCATCATCCCGGTGACGTCTCTCGGTGACCTCGGTGACCTCGGTGACCTCGGTG 420
Db 361 ACGGGCATCATCCCGGTGACGTCTCTCGGTGACCTCGGTGACCTCGGTGACCTCGGTG 420
QY 421 GTACAGGAGACGTGACATTTGGCGGAGGACCCCGCGGATTCATGTTGATTTCCGGG 480
Db 421 GTACAGGAGACGTGACATTTGGCGGAGGACCCCGCGGATTCATGTTGATTTCCGGG 480
QY 481 CGTTACACACCGAGATCAACTCCGCGAGGATGTACGCGCGCGCGGTTCGCGCTCGCTG 540
Db 481 CGTTACACACCGAGATCAACTCCGCGAGGATGTACGCGCGCGCGGTTCGCGCTCGCTG 540
QY 541 TGGCGCGCGCTCAGATGAGGAGCGGTGGAGTACCTGTTTTCGCGCGCGTCCGGT 600
Db 541 TGGCGCGCGCTCAGATGAGGAGCGGTGGAGTACCTGTTTTCGCGCGCGTCCGGT 600
QY 601 TTCAGTTCGCTGCTTGGGGTCTGACGCGGTGGGTCTGATAGGTTCGTTCGCGCGGTCTGA 660
Db 601 TTCAGTTCGCTGCTTGGGGTCTGACGCGGTGGGTCTGATAGGTTCGTTCGCGCGGTCTGA 660
QY 661 TGGTGGCGCGCGCTTCGCGGTATGTGGGTGATGAGCGTCAACCGCGGCGAGCGCGAGC 720
Db 661 TGGTGGCGCGCGCTTCGCGGTATGTGGGTGATGAGCGTCAACCGCGGCGAGCGCGAGC 720
QY 721 TGACCGCGCGCGCTTCGCGGTGCTGGCGGCTACGAGCGCGGTATGGGTGACGG 780
Db 721 TGACCGCGCGCGCTTCGCGGTGCTGGCGGCTACGAGCGCGGTATGGGTGACGG 780
QY 781 TGCCCCCGCGCGGTGATCCCGAGAACCGTGTGAACTGATTCGTATAGCGACCAACC 840
Db 781 TGCCCCCGCGCGGTGATCCCGAGAACCGTGTGAACTGATTCGTATAGCGACCAACC 840
QY 841 TCTTGGGGCAAAACACCCCGCGGATCGCGGTCAACGAGGCGGAAACGCGGAGATGTGG 900
Db 841 TCTTGGGGCAAAACACCCCGCGGATCGCGGTCAACGAGGCGGAAACGCGGAGATGTGG 900
QY 901 CCCAAGACGCGCGCGGTGTTTGGCTACGCGCGCGGACGCGCGAGCGCGAGCGT 960
Db 901 CCCAAGACGCGCGCGGTGTTTGGCTACGCGCGCGGACGCGCGAGCGCGAGCGT 960
QY 961 TGCTGCGGTTCGAGGAGCGCGGAGATGAACGCGCGGTGGGTCTCTCGAGCAGGCGG 1020
Db 961 TGCTGCGGTTCGAGGAGCGCGGAGATGAACGCGCGGTGGGTCTCTCGAGCAGGCGG 1020
QY 1021 CCGCGGTTCGAGGAGCGCTCCGACACCGCGCGCGCAACGAGTTGATGAACATGTGCC 1080

Db 1021 CCGCGTGAAGAGGCTCCGACACCGCGCGCGGCGAACCAGTTGATGAACAATGTGCCCC 1080
QY 1081 AGGCGCTCAACAGCTGGCCAGCCACACGAGGACACACGCCCTTCTTCAAGCTGGGTG 1140
Db 1081 AGGCGCTCAACAGCTGGCCAGCCACACGAGGACACACGCCCTTCTTCAAGCTGGGTG 1140
QY 1141 GCCTGTGAAGACGCTCTCGCGCATCGTTCGCGATCAGTGGTGTGCGATGAGCA 1200
Db 1141 GCCTGTGAAGACGCTCTCGCGCATCGTTCGCGATCAGTGGTGTGCGATGAGCA 1200
QY 1201 ACAACCATGTCGATGACCAACTCGGCTGTGTGATGACCAACACCTTGAGCTCGATG 1260
Db 1201 ACAACCATGTCGATGACCAACTCGGCTGTGTGATGACCAACACCTTGAGCTCGATG 1260
QY 1261 TGAAGGCTTTGCTCCGCGCGCGCGCCGACAGCGCTGTGCAAAACCGCGCGCAAAACGGG 1320
Db 1261 TGAAGGCTTTGCTCCGCGCGCGCGCCGACAGCGCTGTGCAAAACCGCGCGCAAAACGGG 1320
QY 1321 TCCGCGCGATGAGCTCGCTGGGACAGCTCGCTGGGTCTTCTGGGTCTGGGCGTGGGTG 1380
Db 1321 TCCGCGCGATGAGCTCGCTGGGACAGCTCGCTGGGTCTTCTGGGTCTGGGCGTGGGTG 1380
QY 1381 CCGCCAACTTGGTGGCGGCGCTCGGTGGTCTGTTGTCGTCGCCAGCGCTTGGCCG 1440
Db 1381 CCGCCAACTTGGTGGCGGCGCTCGGTGGTCTGTTGTCGTCGCCAGCGCTTGGCCG 1440
QY 1441 CCGCCAACTGAGCTCAACCGCGCGCGCGCTGCGCTGACAGCTGACAGCG 1500
Db 1441 CCGCCAACTGAGCTCAACCGCGCGCGCGCTGCGCTGACAGCTGACAGCG 1500
QY 1501 CCGCGAAAGAGGCGCGGCGAGATGCTGGCGGCGCTCGCGTGGGCGAGATGGCGCCA 1560
Db 1501 CCGCGAAAGAGGCGCGGCGAGATGCTGGCGGCGCTCGCGTGGGCGAGATGGCGCCA 1560
QY 1561 GGGCGGTGGTGGGCTGAGTGGTGTGCTGGTTCGCGCGGACCTATGTGATGCCG 1620
Db 1561 GGGCGGTGGTGGGCTGAGTGGTGTGCTGGTTCGCGCGGACCTATGTGATGCCG 1620
QY 1621 ATTCTCCGCGAGCGCGGATATCCCGCGCGCTTGTGCGAGGACCGGTTCCGCGACT 1680
Db 1621 ATTCTCCGCGAGCGCGGATATCCCGCGCGCTTGTGCGAGGACCGGTTCCGCGACT 1680
QY 1681 TCCCGGCTGCGCTCGACCGTCCGAGTGTGCGGCAAGTGGGCGCCACAGTGGTCA 1740
Db 1681 TCCCGGCTGCGCTCGACCGTCCGAGTGTGCGGCAAGTGGGCGCCACAGTGGTCA 1740
QY 1741 ACATCAACCAACTGGCTTACAAACCGCTGGCGCGCGGCGGCGGATCGTCATCG 1800
Db 1741 ACATCAACCAACTGGCTTACAAACCGCTGGCGCGCGGCGGCGGATCGTCATCG 1800
QY 1801 ATCCCAACGCTGTGCTCAACCAACCAACAGTGTGCGGCGCGGCGGCGGATCAATG 1860
Db 1801 ATCCCAACGCTGTGCTCAACCAACCAACAGTGTGCGGCGCGGCGGCGGATCAATG 1860
QY 1861 CGTTCAGCTCGCTCGCGGCAACCTACCGCTCGATGCTGGTGGGATGACCGCACCC 1920
Db 1861 CGTTCAGCTCGCTCGCGGCAACCTACCGCTCGATGCTGGTGGGATGACCGCACCC 1920
QY 1921 AGGATGTCGGGTGCTGACGCTGGCGGTGCGGCTGCGCTGCGGCGGCGGATCGGTG 1980
Db 1921 AGGATGTCGGGTGCTGACGCTGGCGGTGCGGCTGCGCTGCGGCGGCGGATCGGTG 1980
QY 1981 GCGCGCTCGGCTGTGACCGCTGTGCGGATGGGCAACAGCGTGGGCGGCGGAA 2040
Db 1981 GCGCGCTCGGCTGTGACCGCTGTGCGGATGGGCAACAGCGTGGGCGGCGGAA 2040
QY 2041 CGCCCGTGGGTGCTGGCGAGGCTGTGCGGCTCGGCGCAAAACCGTGCAGGCTCGATT 2100
Db 2041 CGCCCGTGGGTGCTGGCGAGGCTGTGCGGCTCGGCGCAAAACCGTGCAGGCTCGATT 2100
QY 2101 CGTTCAGCTGCGGAGAGACATTGAACGGGTGATGATGCGGCTGATGCGGATCCAGC 2160
Db 2101 CGTTCAGCTGCGGAGAGACATTGAACGGGTGATGATGCGGCTGATGCGGATCCAGC 2160

QY 2161 CCGGTGATTCGGCGCGCGCGCTCGTCAACCGCTAGGACAGGTGGTTCGATGAACACGG 2220
Db 2161 CCGGTGATTCGGCGCGCGCGCTCGTCAACCGCTAGGACAGGTGGTTCGATGAACACGG 2220
QY 2221 CCGGTCTTAGATATCCATCACTGCGCGCGCTCGAGCAGATCCGNTGTAAACAAG 2280
Db 2221 CCGGTCTTAGATATCCATCACTGCGCGCGCTCGAGCAGATCCGNTGTAAACAAG 2280
QY 2281 CCGGAAA 2287
Db 2281 CCGGAAA 2287

RESULT 3

US-10-359-460-1
; Sequence 1, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Devin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein Ra12-TbH9-Ra35 (designated MtB32-Mtb39
; OTHER INFORMATION: fusion)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
US-10-359-460-1

Query Match 99.9%; Score 2284; DB 15; Length 2287;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1	QY	1	CTAGAAATAATTTTGTTTACTTTTAAAGANGANATATACATATNGCATCACCAATCACCATC	60
1	Db	1	CTAGAAATAATTTTGTTTACTTTTAAAGANGANATATACATATNGCATCACCAATCACCATC	60
61	QY	61	ACACGGCCGGCTCCGATTAACCTTCCAGCTGCTCCACGGTGGGCAGAGATTCGCAATTCGGA	120
61	Db	61	ACACGGCCGGCTCCGATTAACCTTCCAGCTGCTCCACGGTGGGCAGAGATTCGCAATTCGGA	120
121	QY	121	TCGGGCAGCGATGGCGATTCGGGGCCAGATCCGATTCGGTGGGGGTCAOCCACCGTTC	180
121	Db	121	TCGGGCAGCGATGGCGATTCGGGGCCAGATCCGATTCGGTGGGGGTCAOCCACCGTTC	180
181	QY	181	ATATCGGGCTACCGCTTTCCTCGCTTGGGTGTTGTCACAAACACACGCAACGCGCGCAC	240
181	Db	181	ATATCGGGCTACCGCTTTCCTCGCTTGGGTGTTGTCACAAACACACGCAACGCGCGCAC	240
241	QY	241	GAGTCCAAACGGCTGGTTCGGGAGCGCTCCGGCGGCAAGTCTCGGCATCTCCACCGCGAAG	300
241	Db	241	GAGTCCAAACGGCTGGTTCGGGAGCGCTCCGGCGGCAAGTCTCGGCATCTCCACCGCGAAG	300
301	QY	301	TGATCACCGGGTTCGACGGCGTCCGATCAATCCGGCCACCGCGATGGCGGACGGCGTTA	360
301	Db	301	TGATCACCGGGTTCGACGGCGTCCGATCAATCCGGCCACCGCGATGGCGGACGGCGTTA	360
361	QY	361	ACGGGCATCATCCGGTGCAGCTCATCTCGGTGACCTTGGCAAAACCAAGTCGGCGGCACGC	420
361	Db	361	ACGGGCATCATCCGGTGCAGCTCATCTCGGTGACCTTGGCAAAACCAAGTCGGCGGCACGC	420
421	QY	421	GTACAGGGAAGTGAATTTGGCCGAGGAGACCCCGGCCGAATTCATGTTGATTTCCGGG	480
421	Db	421	GTACAGGGAAGTGAATTTGGCCGAGGAGACCCCGGCCGAATTCATGTTGATTTCCGGG	480
481	QY	481	CGTTACACCGGAGATCAACTCCGCGAGGATGTAACCCGGCCGGGTTCCGGCTCGCTCG	540
481	Db	481	CGTTACACCGGAGATCAACTCCGCGAGGATGTAACCCGGCCGGGTTCCGGCTCGCTCG	540
541	QY	541	TGSCCCGGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGTTTTTCGGCGCGTCTGGCGT	600
541	Db	541	TGSCCCGGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGTTTTTCGGCGCGTCTGGCGT	600
601	QY	601	TTCAAGTCGGTGGTCTGGGCTCTACCGTGGGTCGTGGATAGTTCTGTCGGCGGGTCTGA	660
601	Db	601	TTCAAGTCGGTGGTCTGGGCTCTACCGTGGGTCGTGGATAGTTCTGTCGGCGGGTCTGA	660
661	QY	661	TGTTGGCGGGCGCTCGCCGTATGTGCGTGGATGAGCGTCAACCGGGGCGAGCGGAGC	720
661	Db	661	TGTTGGCGGGCGCTCGCCGTATGTGCGTGGATGAGCGTCAACCGGGGCGAGCGGAGC	720
721	QY	721	TGACCGCGCCGAGGTCGGGTTCTGCGCGCGCTACGAGACGGGATGGGCTGACCG	780
721	Db	721	TGACCGCGCCGAGGTCGGGTTCTGCGCGCGCTACGAGACGGGATGGGCTGACCG	780
781	QY	781	TGCCCCCGCGGTGATCCCGAGAACCGTCTGAACTGATGTTCTGATAGCCACCAACC	840
781	Db	781	TGCCCCCGCGGTGATCCCGAGAACCGTCTGAACTGATGTTCTGATAGCCACCAACC	840
841	QY	841	TCATTGGGGGAAAACACCCCGCGGATTCGGTCAACGAGGCGGATACGGCGGATGTCG	900
841	Db	841	TCATTGGGGGAAAACACCCCGCGGATTCGGTCAACGAGGCGGATACGGCGGATGTCG	900
901	QY	901	CCCAAGACGCGCGCGATGTTTGGCTACGCGCGGGGACGGGACCGCGCACCGCGCACGT	960
901	Db	901	CCCAAGACGCGCGCGATGTTTGGCTACGCGCGGGGACGGGACCGCGCACCGCGCACGT	960
961	QY	961	TGTTGCCGTTTCAGGACGCGCGGAGATGACACCGCGGGTGGGCTCTCTGAGGAGCGCG	1020
961	Db	961	TGTTGCCGTTTCAGGACGCGCGGAGATGACACCGCGGGTGGGCTCTCTGAGGAGCGCG	1020
1021	QY	1021	CCCGGTCGAGGAGGCGCTCCGACACCGCCGCGGGAACCAAGTTGATCAAAATGTGCCCC	1080
1021	Db	1021	CCCGGTCGAGGAGGCGCTCCGACACCGCCGCGGGAACCAAGTTGATCAAAATGTGCCCC	1080
1081	QY	1081	AGGCGCTGCAACAGCTGCCCGCAGCCACCGAGGCGACCAACGCTTCTTCAAGCTGGGT	1140

	D _b	1081	AGCGCTGCAACAGCTGGCCACCACCGCAGGGCACAGGCCTTCCTTCCAAGCTGGGTG	11440
	Q _y	1141	GCGTGTGAAGACGGTGCTGCCGCATCGGTGCGCATACAGCAAATGTTGTCGATGSCCA	1200
	D _b	1141	GCGTGTGAAGACGGTGCTGCCGCATCGGTGCGCATACAGCAAATGTTGTCGATGSCCA	1200
	Q _y	1201	ACAACACATGTCGATGACCAACTCGGGTGTGTCGATGACCAACACTTGAGCTCGATGT	1260
	D _b	1201	ACAACACATGTCGATGACCAACTCGGGTGTGTCGATGACCAACACTTGAGCTCGATGT	1260
	Q _y	1261	TGAAGGCGCTTGCTCCGGCGCGCGCCGCGCAGGCCGTGCMAAACCGCGGCGCAAAACGGGG	1320
	D _b	1261	TGAAGGCGCTTGCTCCGGCGCGCGCCGCGCAGGCCGTGCMAAACCGCGGCGCAAAACGGGG	1320
	Q _y	1321	TCCGGGCGATGAGCTCGCTGGGACGCTCCGCTGAGTTCTTCGGTCTTGGGCGGTGGGTGG	1380
	D _b	1321	TCCGGGCGATGAGCTCGCTGGGACGCTCCGCTGAGTTCTTCGGTCTTGGGCGGTGGGTGG	1380
	Q _y	1381	C CGCCAACTTTGGGTGCGGCGCGGCGCTCGGTTCGTTGTTCGGTCCGCGACGCTTGCGCG	1440
	D _b	1381	C CGCCAACTTTGGGTGCGGCGCGGCGCTCGGTTCGTTGTTCGGTCCGCGACGCTTGCGCG	1440
	Q _y	1441	CGGCCAACAGGACAGTCACCCC GGCGCGCGGGCGCTGCCGTGACAGGCTGACACAGC	1500
	D _b	1441	CGGCCAACAGGACAGTCACCCC GGCGCGCGGGCGCTGCCGTGACAGGCTGACACAGC	1500
	Q _y	1501	CCGCGGAAAGAGGSCCGCGGACAGATCTGGCGGGCTCGCGGTGGGCGAGATGGGCGCCA	1560
	D _b	1501	CCGCGGAAAGAGGSCCGCGGACAGATCTGGCGGGCTCGCGGTGGGCGAGATGGGCGCCA	1560
	Q _y	1561	GGGCGGTTGGTGGGCTCAGTGGTGTGTGGTGTTCGGCGCGACCTATGTGATGCCGC	1620
	D _b	1561	GGGCGGTTGGTGGGCTCAGTGGTGTGTGGTGTTCGGCGCGACCTATGTGATGCCGC	1620
	Q _y	1621	ATTCTCGGCGACCGCGCGATATCGCCCGCGCGCCTTGTGCGAGGACCGGTTCCGCCACT	1680
	D _b	1621	ATTCTCGGCGACCGCGCGATATCGCCCGCGCGCCTTGTGCGAGGACCGGTTCCGCCACT	1680
	Q _y	1681	TCCCGCGCTCCCCCTCGACCCGTCGCGATGTGCGCCAAAGTGGGGCCACAGTGTGTCFA	1740
	D _b	1681	TCCCGCGCTCCCCCTCGACCCGTCGCGATGTGCGCCAAAGTGGGGCCACAGTGTGTCFA	1740
	Q _y	1741	ACATCAACACAAACTGGGCTAACAACCGCGTGGGCGCGGACCGGCTATCGTCAATCG	1800
	D _b	1741	ACATCAACACAAACTGGGCTAACAACCGCGTGGGCGCGGACCGGCTATCGTCAATCG	1800
	Q _y	1801	ATCCCAAGGTGTGTTGTGTACCAACAACACACGTGATCGCGGGCGCCACCCACATCAATG	1860
	D _b	1801	ATCCCAAGGTGTGTTGTGTACCAACAACACACGTGATCGCGGGCGCCACCCACATCAATG	1860
	Q _y	1861	CGTTACGCGTCCGGTCCGGCCAAACCTACGGCGTCGATGTGGTTCGGGTATGACCGCACCC	1920
	D _b	1861	CGTTACGCGTCCGGTCCGGCCAAACCTACGGCGTCGATGTGGTTCGGGTATGACCGCACCC	1920
	Q _y	1921	AGGATGTGCGGTGTGAGCTGCGGTGCGCGTGGCGCTCGCTGCGCGCGATCGCGTG	1980
	D _b	1921	AGGATGTGCGGTGTGAGCTGCGGTGCGCGTGGCGCTCGCTGCGCGCGATCGCGTG	1980
	Q _y	1981	GCGGCGTCCGCTTGGTCAGCCCGCTGTCGCGGATGGGCAACAGCGGTGGGCGAGCGCGAA	2040
	D _b	1981	GCGGCGTCCGCTTGGTCAGCCCGCTGTCGCGGATGGGCAACAGCGGTGGGCGAGCGCGAA	2040
	Q _y	2041	GCGCCCGTCCGCTGCGTAGGTTGTGCGCTCGCGCCAAACCGTGCAGGGTTCGGATT	2100
	D _b	2041	GCGCCCGTCCGCTGCGTAGGTTGTGCGCTCGCGCCAAACCGTGCAGGGTTCGGATT	2100
	Q _y	2101	CGCTACCGGTGCGGAGAGACATTGAACCGGTTGATCCAGTTCCGATCCGCGATCCAGC	2160
	D _b	2101	CGCTACCGGTGCGGAGAGACATTGAACCGGTTGATCCAGTTCCGATCCGCGATCCAGC	2160
	Q _y	2161	CCGCGTGAATTCGGGCGGGCCCGCTGCTCAACCGGCTTAGGACAGGTGTGCGGTATGAACACGG	2220

Db 2161 CCGGTGATTCGGCGGGCCCGCTCGTCAACGGCTAGACAGGTGGTATGAACACGG 2220
 Qy 2221 CCGCGTCTAGGATATCATCAGCTGCGCGCGCTCGAGCAGATCCGNTGTAAACAAG 2280
 Db 2221 CCGCGTCTAGGATATCCATCAGCTGCGCGCGCTCGAGCAGATCCGNTGTAAACAAG 2280
 Qy 2281 CCGGAAA 2287
 Db 2281 CCGGAAA 2287
 RESULT 4
 US-10-098-732A-15
 ; Sequence 15, Application US/10098732A
 ; Publication No. US20030175294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Brannon, Mark
 ; APPLICANT: Guderian, Jeffrey
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; FILE OF INVENTION: Leishmania Antigen
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: US 60/275,837
 ; PRIOR FILING DATE: 2001-03-13
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 15
 ; LENGTH: 2287
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
 ; OTHER INFORMATION: protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
 ; OTHER INFORMATION: fusion)
 ; NAME/KEY: CDS
 ; LOCATION: (42)..(2231)
 ; OTHER INFORMATION: MTB72F
 ; NAME/KEY: modified base
 ; LOCATION: (1)..(2287)
 ; OTHER INFORMATION: n = g, a, c or t
 US-10-098-732A-15
 Query Match 99.9%; Score 2284; DB 15; Length 2287;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TCTAGAATAATTTGTTTACTTTAGAANGANATATACATATGCATCACCATCACCATC 60
 Db 1 TCTAGAATAATTTGTTTACTTTAGAANGANATATACATATGCATCACCATCACCATC 60
 Qy 61 ACAGGCGCGTCCGATAAATTCAGCTGTCCAGAGGTGGGAGGATTCGCCATTCGGA 120
 Db 61 ACAGGCGCGTCCGATAAATTCAGCTGTCCAGAGGTGGGAGGATTCGCCATTCGGA 120
 Qy 121 TCGGCGAGGCGATGGCGATCGCGGCGAGATCCGATCGGTTGGGTTGGGTTGGGTTGG 180
 Db 121 TCGGCGAGGCGATGGCGATCGCGGCGAGATCCGATCGGTTGGGTTGGGTTGGGTTGG 180
 Qy 181 ATATCGGCGCTACCGCTTCTCGGTTGGGTTGGTTCGACAAACACGCGCAACGCGGCAC 240
 Db 181 ATATCGGCGCTACCGCTTCTCGGTTGGGTTGGTTCGACAAACACGCGCAACGCGGCAC 240
 Qy 241 GAGTCCAAACGCTGGTGGAGCGCTCCGCGCGCAAGTCTCGGATCTCCACCGCGAGC 300
 Db 241 GAGTCCAAACGCTGGTGGAGCGCTCCGCGCGCAAGTCTCGGATCTCCACCGCGAGC 300
 Qy 301 TGATCAACCGGCTCGACGGGCTCCGATCAACTCGGCGCACCGGATGGCGGACGCGCTTA 360

Db 301 TGATCAACCGGCTCGACGGCTCCGATCAACTCGGCGCACCGGATGGCGGACGCGCTTA 360
 Qy 361 ACGGCGATCATCCGCTGACGTCTCTCGGTGACCTTGGCAACCAAGTCGGCGGCGACGC 420
 Db 361 ACGGCGATCATCCGCTGACGTCTCTCGGTGACCTTGGCAACCAAGTCGGCGGCGACGC 420
 Qy 421 GTACAGGGAACGTGACATTTGGCCGAGGAGACCCCGCGCGCAATTCATGGTGGATTTCCGGG 480
 Db 421 GTACAGGGAACGTGACATTTGGCCGAGGAGACCCCGCGCGCAATTCATGGTGGATTTCCGGG 480
 Qy 481 CGTTACACCGGAGATCAACTCCGCGAGGATGTACGCGCGCGCGGTTTCGGCTTCGCTGG 540
 Db 481 CGTTACACCGGAGATCAACTCCGCGAGGATGTACGCGCGCGCGGTTTCGGCTTCGCTGG 540
 Qy 541 TGGCGCGCGCTCAGATGTGGGACACGCTGGCGAGTGACCTGTTTCGCGCGCTCGCGGT 600
 Db 541 TGGCGCGCGCTCAGATGTGGGACACGCTGGCGAGTGACCTGTTTCGCGCGCTCGCGGT 600
 Qy 601 TTCAGTCCGTGCTTCGGGCTCTGACCGTGGGCTCGTGCATAGGTTCTTCGGCGGTCTGA 660
 Db 601 TTCAGTCCGTGCTTCGGGCTCTGACCGTGGGCTCGTGCATAGGTTCTTCGGCGGTCTGA 660
 Qy 661 TGGTGGCGCGCGCTCCGCTATGTGGCGTGAGGCTCACCGCGCGGCGAGCGCGAGC 720
 Db 661 TGGTGGCGCGCGCTCCGCTATGTGGCGTGAGGCTCACCGCGCGGCGAGCGCGAGC 720
 Qy 721 TGACCGCGCGCGCGCTCCGCGTTCGCGCGCGCGCTACGAGACGCGCTATTCGGGTGACGG 780
 Db 721 TGACCGCGCGCGCGCTCCGCGTTCGCGCGCGCGCTACGAGACGCGCTATTCGGGTGACGG 780
 Qy 781 TGCGCGCGCGCGCTTCGCGAGAACCGTGTGACATGATGATTCGATAGCAGCACACC 840
 Db 781 TGCGCGCGCGCGCTTCGCGAGAACCGTGTGACATGATGATTCGATAGCAGCACACC 840
 Qy 841 TCTTGGGCGAAAAACACCCCGCGCTCGCGGTCAACAGGCGCGAATACGCGAGATGTGGG 900
 Db 841 TCTTGGGCGAAAAACACCCCGCGCTCGCGGTCAACAGGCGCGAATACGCGAGATGTGGG 900
 Qy 901 CCCAAGACGCGCGCGCTGTTTGGCTACGCGCGCGCGACGCGCGACGCGCGACGT 960
 Db 901 CCCAAGACGCGCGCGCTGTTTGGCTACGCGCGCGCGACGCGCGACGCGCGACGT 960
 Qy 961 TGCTGCGCTTCGAGAGGCGCGCGAGATGACCGCGCGGTTGGCTCTTCGAGCAGCGCG 1020
 Db 961 TGCTGCGCTTCGAGAGGCGCGCGAGATGACCGCGCGGTTGGCTCTTCGAGCAGCGCG 1020
 Qy 1021 CCGCGCTCAGAGGCGCTCCGACACCGCGCGCGAACCAGTTGATGAACAATGTGCCCC 1080
 Db 1021 CCGCGCTCAGAGGCGCTCCGACACCGCGCGCGAACCAGTTGATGAACAATGTGCCCC 1080
 Qy 1081 AGGCGCTGACACAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
 Db 1081 AGGCGCTGACACAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
 Qy 1141 GCCTGTGGAACACGCTTCGCGCGCATCGCTCGCGCGATCAGCAACATCGTTCGATGCGCCA 1200
 Db 1141 GCCTGTGGAACACGCTTCGCGCGCATCGCTCGCGCGATCAGCAACATCGTTCGATGCGCCA 1200
 Qy 1201 ACAACACATGTTCGATGACCACTCGGCTGTGTCGATGACCAACCTTCGATGCTGATGT 1260
 Db 1201 ACAACACATGTTCGATGACCACTCGGCTGTGTCGATGACCAACCTTCGATGCTGATGT 1260
 Qy 1261 TGAAGGCTTTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
 Db 1261 TGAAGGCTTTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
 Qy 1321 TCGGCGGATGAGCTCGCTCGGCGAGCTCGCTGGGTTCTTCGGGTTCTTCGGGTTGGGTTGG 1380
 Db 1321 TCGGCGGATGAGCTCGCTCGGCGAGCTCGCTGGGTTCTTCGGGTTCTTCGGGTTGGGTTGG 1380
 Qy 1381 CCGCAACTTTGGGTCGGCGCGCGCTTCGCTCGGTTGCTGGTTCGGTTCGGGCGCTGGCGCG 1440
 Db 1381 CCGCAACTTTGGGTCGGCGCGCGCTTCGCTCGGTTGCTGGTTCGGTTCGGGCGCTGGCGCG 1440

Db 541 TGGCCGCGCTCAGATGTGGGACACGGTGGGAGTGAACCTGTGTTTCGCGCGCTCGGCGT 600
QY 601 TTCAGTCCGTGCTCTGGGCTCTGACCGGTGGGTCTGTGATAGGTTCGTGCGCGGTCTGA 660
Db 601 TTCAGTCCGTGCTCTGGGCTCTGACCGGTGGGTCTGTGATAGGTTCGTGCGCGGTCTGA 660
QY 661 TGGTGGCGCGCCCTCGCCGTAATGTGGCGTGAATGAGCGTCAACCGCGGGGAGCGCGAGC 720
Db 661 TGGTGGCGCGCCCTCGCCGTAATGTGGCGTGAATGAGCGTCAACCGCGGGGAGCGCGAGC 720
QY 721 TGACCGCGCCCGAGGTTCGGGTTGCTCGCGGGCTACGACGCGGTATGGGTGACGG 780
Db 721 TGACCGCGCCCGAGGTTCGGGTTGCTCGCGGGCTACGACGCGGTATGGGTGACGG 780
QY 781 TGCCCGCGCGGTGATCGCCGAGAACCGTGTGAATCTGATAGCGACCAACC 840
Db 781 TGCCCGCGCGGTGATCGCCGAGAACCGTGTGAATCTGATAGCGACCAACC 840
QY 841 TCTTGGGGCAAAACACCCCGGCTCGCGGTCAACGAGCGCAATAGCGGAGATGGG 900
Db 841 TCTTGGGGCAAAACACCCCGGCTCGCGGTCAACGAGCGCAATAGCGGAGATGGG 900
QY 901 CCCAAGAGCGCCCGCGATGTTGGCTACGCGCGGGCGACCGCGCGCGACGCGCGACGT 960
Db 901 CCCAAGAGCGCCCGCGATGTTGGCTACGCGCGGGCGACCGCGCGCGACGCGCGACGT 960
QY 961 TGCTGCGGTCGAGGAGCGCGGAGATGACACGCGCGGGTGGGCTCTCTGAGAGCGCG 1020
Db 961 TGCTGCGGTCGAGGAGCGCGGAGATGACACGCGCGGGTGGGCTCTCTGAGAGCGCG 1020
QY 1021 CCGCGCTCGAGAGCGCTCCGACACCGCGCGGGCAACCACTGATCAACATGTGCCCC 1080
Db 1021 CCGCGCTCGAGAGCGCTCCGACACCGCGCGGGCAACCACTGATCAACATGTGCCCC 1080
QY 1081 AGGCGCTCAACAGTGGCGCGCCAGCCGACGAGGCGCAACGCGCTTTCGAGCTGGGT 1140
Db 1081 AGGCGCTCAACAGTGGCGCGCCAGCCGACGAGGCGCAACGCGCTTTCGAGCTGGGT 1140
QY 1141 GCCTGTGGAAGACGCTTCGCGCGATCGCTGCGCGATCAGCACATGTCGATGCGCA 1200
Db 1141 GCCTGTGGAAGACGCTTCGCGCGATCGCTGCGCGATCAGCACATGTCGATGCGCA 1200
QY 1201 ACAACCAATGTCGATGACCAACTCGGCTGTGCTGATGACCAACACCTTGAGCTCGATG 1260
Db 1201 ACAACCAATGTCGATGACCAACTCGGCTGTGCTGATGACCAACACCTTGAGCTCGATG 1260
QY 1261 TGAAGGGCTTTGCTCG 1320
Db 1261 TGAAGGGCTTTGCTCG 1320
QY 1321 TCCGGGCGATGAGCTCGCTGGGAGCTCGCTGGGTTCTTCGGGTCTGGGCGGTGGGTTG 1380
Db 1321 TCCGGGCGATGAGCTCGCTGGGAGCTCGCTGGGTTCTTCGGGTCTGGGCGGTGGGTTG 1380
QY 1381 CCGGCACTTGGGTTCGGGCGCGCTCGCTGGGTTCTTCGGGTCTGGGCGGTGGGTTG 1440
Db 1381 CCGGCACTTGGGTTCGGGCGCGCTCGCTGGGTTCTTCGGGTCTGGGCGGTGGGTTG 1440
QY 1441 CCGGCACTTGGGTTCGGGCGCGCTCGCTGGGTTCTTCGGGTCTGGGCGGTGGGTTG 1500
Db 1441 CCGGCACTTGGGTTCGGGCGCGCTCGCTGGGTTCTTCGGGTCTGGGCGGTGGGTTG 1500
QY 1501 CCGGCACTTGGGTTCGGGCGCGCTCGCTGGGTTCTTCGGGTCTGGGCGGTGGGTTG 1560
Db 1501 CCGGCACTTGGGTTCGGGCGCGCTCGCTGGGTTCTTCGGGTCTGGGCGGTGGGTTG 1560
QY 1561 GGGCGGCTGGTGGCTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1620
Db 1561 GGGCGGCTGGTGGCTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1620
QY 1621 ATTCTCCGAGCGCGGATATCGCCCGCGCGCTTGTGCGAGGACCGGTTTCGCGACT 1680
Db 1621 ATTCTCCGAGCGCGGATATCGCCCGCGCGCTTGTGCGAGGACCGGTTTCGCGACT 1680

QY 1681 TCCCCCGCTGCGCTCGACCGGTTCGCGATGCTGCCCAAGTGGGGCCACAGGTGGTCA 1740
Db 1681 TCCCCCGCTGCGCTCGACCGGTTCGCGATGCTGCCCAAGTGGGGCCACAGGTGGTCA 1740
QY 1741 ACATCAACCAAACTGGGCTACAACACCGCGTGGGCGCCCGGACCGGATCGTCATCG 1800
Db 1741 ACATCAACCAAACTGGGCTACAACACCGCGTGGGCGCCCGGACCGGATCGTCATCG 1800
QY 1801 ATCCCAACCGTTCGCTGCTGACCAACACCGTTCGCGGGCGCCACCGATCAATG 1860
Db 1801 ATCCCAACCGTTCGCTGCTGACCAACACCGTTCGCGGGCGCCACCGATCAATG 1860
QY 1861 CGTTACAGCGTTCGCTGCTGCGGCGCAAACTACGCGGTTCGATGCTGGGTATGACCGCACCC 1920
Db 1861 CGTTACAGCGTTCGCTGCTGCGGCGCAAACTACGCGGTTCGATGCTGGGTATGACCGCACCC 1920
QY 1921 AGGATGTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 1980
Db 1921 AGGATGTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 1980
QY 1981 GCGGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2040
Db 1981 GCGGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2040
QY 2041 CCGCGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2100
Db 2041 CCGCGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2100
QY 2101 CCGTTCGCGGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2160
Db 2101 CCGTTCGCGGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2160
QY 2161 CCGGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2220
Db 2161 CCGGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2220
QY 2221 CCGGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2280
Db 2221 CCGGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2280
QY 2281 CCGGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2340
Db 2281 CCGGCGTTCGCGGTTCGCTGCTGCGGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 2340

RESULT 6

US-10-369-983-8
; Sequence 8, Application US/10369983
; Publication No. US2003023593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB81F (MTB72F-DPV)
US-10-369-983-8

Query Match 95.6%; Score 2186.8; DB 16; Length 2451;

Best Local Similarity 99.9%; Pred. No. 0; Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	39	CATATGCATCACCATCACCATCACAGCGCGGTCCGATACCTCCAGCTGCCAGAGT	98
Db	1	CATATGCATCACCATCACCATCACAGCGCGGTCCGATACCTCCAGCTGCCAGAGT	60
QY	99	GGCGAGGATTCGCCATTCGGATCGGCGAGGCGATCGCGGCGAGATCCGATCGATCG	158
Db	61	GGCGAGGATTCGCCATTCGGATCGGCGAGGCGATCGCGGCGAGATCCGATCGATCG	120
QY	159	GGTGGGGGTACCCACCGTTCATATCGGGGCTACCGGCTTCCTGGCTGGGTTGTC	218
Db	121	GGTGGGGGTACCCACCGTTCATATCGGGGCTACCGGCTTCCTGGCTGGGTTGTC	180
QY	219	GACAAACAGCGGCAAGCGGCGAGTCCAAACGGTTCGGAGGCGCTCCGCGCAAGT	278
Db	181	GACAAACAGCGGCAAGCGGCGAGTCCAAACGGTTCGGAGGCGCTCCGCGCAAGT	240
QY	279	CTCGGATCTCCACCGGCGAGCTGATCACCGGTCGACGGGCTCCGATCAACTCGGCC	338
Db	241	CTCGGATCTCCACCGGCGAGCTGATCACCGGTCGACGGGCTCCGATCAACTCGGCC	300
QY	339	ACGCGATGGCGAGCGCTTAACGGGCTATCCGGTGAGCTCATCTCGGTGACCTGG	398
Db	301	ACGCGATGGCGAGCGCTTAACGGGCTATCCGGTGAGCTCATCTCGGTGACCTGG	360
QY	399	CAACCAAGTCGGCGGCGCGCTACAGGGAAGTGCATTCGCGGAGGAGCCCGGCC	458
Db	361	CAACCAAGTCGGCGGCGCGCTACAGGGAAGTGCATTCGCGGAGGAGCCCGGCC	420
QY	459	GAATTCATGTGATTTGGGGGCTTACACCGGAGATCAATCCCGGAGGATGTACGCC	518
Db	421	GAATTCATGTGATTTGGGGGCTTACACCGGAGATCAATCCCGGAGGATGTACGCC	480
QY	519	GGCGCGGTTGGCTTCGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGGCTGAC	578
Db	481	GGCGCGGTTGGCTTCGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGGCTGAC	540
QY	579	CTGTTTCGCGCGCGTGGCTTCAGTGGTTCGGTTCGGTTCGAGCGTGGGCTGTCG	638
Db	541	CTGTTTCGCGCGCGTGGCTTCAGTGGTTCGGTTCGGTTCGAGCGTGGGCTGTCG	600
QY	639	ATAGGTTCTGGCGGGTCTGATGGTGGCGGCTCCCGCTATCTGGCTGGATGACG	698
Db	601	ATAGGTTCTGGCGGGTCTGATGGTGGCGGCTCCCGCTATCTGGCTGGATGACG	660
QY	699	GTCAACCGCGGCGAGGCTGAGCGCGCGGCTCGGCTCGGCTTCGCGGCGCTAC	758
Db	661	GTCAACCGCGGCGAGGCTGAGCGCGCGGCTCGGCTCGGCTTCGCGGCGCTAC	720
QY	759	GAGACGCGTATGGCTGACGCTGCGCGCGGCTGATCGCGGAGAACCGTGTGAACTG	818
Db	721	GAGACGCGTATGGCTGACGCTGCGCGCGGCTGATCGCGGAGAACCGTGTGAACTG	780
QY	819	ATGATTCGATAGCGCAACCTCTTGGGCAAAACACCGCGCGATCGCGGTCAACGAG	878
Db	781	ATGATTCGATAGCGCAACCTCTTGGGCAAAACACCGCGCGATCGCGGTCAACGAG	840
QY	879	GCCGAATACCGGAGATGTGGGCGGCAAGACCGCGCGGATTTGGCTACGCGCGCG	938
Db	841	GCCGAATACCGGAGATGTGGGCGGCAAGACCGCGCGGATTTGGCTACGCGCGCG	900
QY	939	ACGCGACGCGGAGCGGAGTTCCTGCTGCTGAGGAGCGCGGAGATGACACGCG	998
Db	901	ACGCGACGCGGAGCGGAGTTCCTGCTGCTGAGGAGCGCGGAGATGACACGCG	960
QY	999	GGTGGGCTCTTCGAGCAGGCGCGCGCTCGAGAGGCTCCGACACCGCGCGGCAAC	1058
Db	961	GGTGGGCTCTTCGAGCAGGCGCGCGCTCGAGAGGCTCCGACACCGCGCGGCAAC	1020
QY	1059	CAGTTGATGAACAATGTCCCGCAGCGCTGCAACAGCTGGCGGCGGCGGCGGCGG	1118
Db	1021	CAGTTGATGAACAATGTCCCGCAGCGCTGCAACAGCTGGCGGCGGCGGCGGCGG	1080
QY	1119	ACGCTTCTTCCAGCTGGGTGGCTGTGGAAGACGGTCTTCGCGCATCGGTCCCGATC	1178
Db	1081	ACGCTTCTTCCAGCTGGGTGGCTGTGGAAGACGGTCTTCGCGCATCGGTCCCGATC	1140
QY	1179	AGCAACATGTGTGATGGCGCAACCAACATGTGATGACCAACTCGGCTGTGTCATG	1238
Db	1141	AGCAACATGTGTGATGGCGCAACCAACATGTGATGACCAACTCGGCTGTGTCATG	1200
QY	1239	ACCAACACCTTGAATCGATGTTGAAGGGCTTTGCTCCGGCGGCGGCGGCGGCGG	1298
Db	1201	ACCAACACCTTGAATCGATGTTGAAGGGCTTTGCTCCGGCGGCGGCGGCGGCGG	1260
QY	1299	CAAAACCGCGGCAAAACCGGGTCCGGCGGATGAGCTCGCTGGGACGCTCGCTGGTCT	1358
Db	1261	CAAAACCGCGGCAAAACCGGGTCCGGCGGATGAGCTCGCTGGGACGCTCGCTGGTCT	1320
QY	1359	TCGGTCTGGCGGTGGGTGGCGGCAACTTGGGTGGCGGCTCGGTGGTTCGTTG	1418
Db	1321	TCGGTCTGGCGGTGGGTGGCGGCAACTTGGGTGGCGGCTCGGTGGTTCGTTG	1380
QY	1419	TCGGTGGCGGAGTGGCGGCGGCGGCAACAGGAGTCAACCGCGGCGGCGGCGGCTG	1478
Db	1381	TCGGTGGCGGAGTGGCGGCGGCGGCAACAGGAGTCAACCGCGGCGGCGGCGGCTG	1440
QY	1479	CGCTGACGAGCTGACAGCGCGCGGAGAGGCGCGGCGGAGATGCTGGGCGGCTG	1538
Db	1441	CGCTGACGAGCTGACAGCGCGCGGAGAGGCGCGGCGGAGATGCTGGGCGGCTG	1500
QY	1539	CGGTGGGCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG	1598
Db	1501	CGGTGGGCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG	1560
QY	1599	CGGCAACCTATGTGATGCGCGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCTG	1658
Db	1561	CGGCAACCTATGTGATGCGCGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCTG	1620
QY	1659	TCGAGGAGCGGTTCCCGGCTTCGCGGCTTCGAGCGGCTCGCGGCTCGCGGCTCGCC	1718
Db	1621	TCGAGGAGCGGTTCCCGGCTTCGCGGCTTCGAGCGGCTCGCGGCTCGCGGCTCGCC	1680
QY	1719	CAAGTGGGCGCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACCGCGTGGG	1778
Db	1681	CAAGTGGGCGCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACCGCGTGGG	1740
QY	1779	GCGGAGACGCGATCGTATCGATCCGAGCGGTGCTGCTGAGCAACCAACAGCGTATC	1838
Db	1741	GCGGAGACGCGATCGTATCGATCCGAGCGGTGCTGCTGAGCAACCAACAGCGTATC	1800
QY	1839	GCGGCGCACCGACATCAATGCGTTCAAGCGTTCGCGCTCCGCGCAAACTACGCGTCCG	1898
Db	1801	GCGGCGCACCGACATCAATGCGTTCAAGCGTTCGCGCTCCGCGCAAACTACGCGTCCG	1860
QY	1899	GTGGTGGGATGACCGGACCGGATGTGCGGCTGCTGAGCTGCGCGGTGCGCGTGGC	1958
Db	1861	GTGGTGGGATGACCGGACCGGATGTGCGGCTGCTGAGCTGCGCGGTGCGCGTGGC	1920
QY	1959	CTGCGCTGGCGGCGATCGGTGGCGGCTTCGCGTTGGTGGAGCGGCTCGTCCGATGGG	2018
Db	1921	CTGCGCTGGCGGCGATCGGTGGCGGCTTCGCGTTGGTGGAGCGGCTCGTCCGATGGG	1980
QY	2019	AAACAGCGTGGGAGGCGGAAACCGCGCTGCTGCGGTGCTGGCAGGCTGGTCCGCTCGG	2078
Db	1981	AAACAGCGTGGGAGGCGGAAACCGCGCTGCTGCGGTGCTGGCAGGCTGGTCCGCTCGG	2040
QY	2079	CAAAACGCTGAGGCGGATTCGCTGACCGGTCGAGAGGAGACATTCGAGCGGTGATC	2138
Db	2041	CAAAACGCTGAGGCGGATTCGCTGACCGGTCGAGAGGAGACATTCGAGCGGTGATC	2100
QY	2139	CAGTTGATGCGCGATCCAGCGCGGCTGATTGGGCGGCGGCGGCTCGTCAACGCGCTAG	2198
Db	2101	CAGTTGATGCGCGATCCAGCGCGGCTGATTGGGCGGCGGCGGCTCGTCAACGCGCTAG	2160

```
QY 2199 CAGGTGGTGGTATGAAACACGCGCGCTCC 2228
DB 2161 CAGGTGGTGGTATGAAACACGCGCGCTCC 2190

RESULT 7
US-10-369-983-7
; Sequence 7, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2487
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB83F (MTB72F-MT1)
US-10-369-983-7

Query Match 95.6%; Score 2186.8; DB 16; Length 2487;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 39 CATATGATCAACCATCAACACGCGCGCTCCGATTAACCTCCAGCTGTCCTCCAGGT 98
DB 1 CATATGATCAACCATCAACACGCGCGCTCCGATTAACCTCCAGCTGTCCTCCAGGT 60

QY 99 GGGCAGGATTCGCAATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCG 158
DB 61 GGGCAGGATTCGCAATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCG 120

QY 159 GGTGGGGGTCAACCCAGGTTCATATCGGGCTACCGCTTCTCCGCTTGGGTGTTGC 218
DB 121 GGTGGGGGTCAACCCAGGTTCATATCGGGCTACCGCTTCTCCGCTTGGGTGTTGC 180

QY 219 GACAAACACGCGCAACGCGCACGAGTCCACGCGTGTGTCCGGAGCGCTCCGGCGGCAAGT 278
DB 181 GACAAACACGCGCAACGCGCACGAGTCCACGCGTGTGTCCGGAGCGCTCCGGCGGCAAGT 240

QY 279 CTCGGATCTCCACCGCGCAGCTGATCACCGCGTGTGTCCGGAGCGCTCCGGCGGCAAGT 338
DB 241 CTCGGATCTCCACCGCGCAGCTGATCACCGCGTGTGTCCGGAGCGCTCCGGCGGCAAGT 300

QY 339 ACCCGATGCGGACGCGCTTAACGGGATCATCCCGGTGACGTCTCCGGTGAACCTCG 398
DB 301 ACCCGATGCGGACGCGCTTAACGGGATCATCCCGGTGACGTCTCCGGTGAACCTCG 360

QY 399 CAAACCAAGTCCGGCGGCAACGCTGACAGGAAAGTGTGACATTTGGCGGAGGACCCCGGCC 458
DB 361 CAAACCAAGTCCGGCGGCAACGCTGACAGGAAAGTGTGACATTTGGCGGAGGACCCCGGCC 420

QY 459 GAATTCATGTGATTTCCGGGGGTTTACACCGGAGATCAACTCCCGAGGATGTACGCC 518
DB 421 GAATTCATGTGATTTCCGGGGGTTTACACCGGAGATCAACTCCCGAGGATGTACGCC 480

QY 519 GGCCCGGGTTCGGCTCGCTGTTGGCGCGGCTCAGATGTGGACACGCTGGCGAGTGAC 578
DB 481 GGCCCGGGTTCGGCTCGCTGTTGGCGCGGCTCAGATGTGGACACGCTGGCGAGTGAC 540

QY 579 CTGTTTTCGGCGCGCTGGGTTTTCAGTGGTGGTCTGGGGTCTGAGCGTGGGTGCTGG 638

541 CTGTTTTCGGCGCGCTGGGTTTTCAGTGGTGGTCTGGGGTCTGACCGTGGGTGCTGG 600
639 ATAGGTTCTCGCGCGGTCTGATGTTGGCGCGGCTCTCGCGTATGTTGGCGGTGAGTAC 698
601 ATAGGTTCTCGCGCGGTCTGATGTTGGCGCGGCTCTCGCGTATGTTGGCGGTGAGTAC 660
699 GTACCGCGCGGCGAGCGGTGACCGCGCGCGGTGACCGCGCGGTGACCGCGCGGTGAC 758
661 GTACCGCGCGGCGAGCGGTGACCGCGCGCGGTGACCGCGCGGTGACCGCGCGGTGAC 720
759 GAGCGCGGTATGGGTGACCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGAC 818
721 GAGCGCGGTATGGGTGACCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGAC 780
819 ATGATTTCTGATAGCGACCAACCTCTTGGGCAAAACACCGCGGTGACCGGTGACCGAG 878
781 ATGATTTCTGATAGCGACCAACCTCTTGGGCAAAACACCGCGGTGACCGGTGACCGAG 840
879 GCGGATACGCGAGATGTGGCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCG 938
841 GCGGATACGCGAGATGTGGCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCG 900
939 ACGGCGCGCGGCGAGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCG 998
901 ACGGCGCGCGGCGAGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCG 960
999 GGTGGGTCTCTGACGCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGAC 1058
961 GGTGGGTCTCTGACGCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGAC 1020
1059 CAGTTGATGAACAATGTGCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGT 1118
1021 CAGTTGATGAACAATGTGCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGT 1080
1119 ACGCCCTCTTCAAGCTGGGTGGCTGTGGAAGACGGTCTCGCGCATCGGTGCGCGCATC 1178
1081 ACGCCCTCTTCAAGCTGGGTGGCTGTGGAAGACGGTCTCGCGCATCGGTGCGCGCATC 1140
1179 ACACAATGGTGTGATGGCGCAACCAACATGTCGATGACCAACTCGGTGTCGATG 1238
1141 ACACAATGGTGTGATGGCGCAACCAACATGTCGATGACCAACTCGGTGTCGATG 1200
1239 ACCACACCTTGAGTCTGATGTCGAGGCTTCTCGCGCGGTGACCGCGCGGTGACCGCGCG 1298
1201 ACCACACCTTGAGTCTGATGTCGAGGCTTCTCGCGCGGTGACCGCGCGGTGACCGCGCG 1260
1299 CAAACCGCGCGCAACCGGGTCTCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCG 1358
1261 CAAACCGCGCGCAACCGGGTCTCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCG 1320
1359 TCGGCTCTGGCGGTGGGTGGCGCGCACTTGGGTGGCGCGGTGACCGCGCGGTGACCGCGCG 1418
1321 TCGGCTCTGGCGGTGGGTGGCGCGCACTTGGGTGGCGCGGTGACCGCGCGGTGACCGCGCG 1380
1419 TCGGCTCTGGCGGTGGGTGGCGCGCACTTGGGTGGCGCGGTGACCGCGCGGTGACCGCGCG 1478
1381 TCGGCTCTGGCGGTGGGTGGCGCGCACTTGGGTGGCGCGGTGACCGCGCGGTGACCGCGCG 1440
1479 CCGTGTACCGCTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCG 1538
1441 CCGTGTACCGCTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCG 1500
1539 CCGTGTGGCGGATGGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCG 1598
1501 CCGTGTGGCGGATGGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCG 1560
1599 CCGCGACCTATGATGCGCGCATTTCTCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCG 1658
1561 CCGCGACCTATGATGCGCGCATTTCTCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCG 1620
1659 TCGCAGGACCGGTGCGCGCATTTCTCGCGCGGTGACCGCGCGGTGACCGCGCGGTGACCGCGCG 1718
```

1621	Db	TCGAGGACGGGTTCCGGCGATTTCCCGGGCTGCCCTCTGACCCGTCGCGATGGTCGCC	1680
1719	QY	CAAGTGGGGCCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACACGCCGTGGGC	1778
1681	Db	CAAGTGGGGCCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACACGCCGTGGGC	1740
1779	QY	GCGGGACCGGCATCGTATCGATCCACAGGTGCTGCTGACCAACACACACGTGATC	1838
1741	Db	GCGGGACCGGCATCGTATCGATCCACAGGTGCTGCTGACCAACACACGTGATC	1800
1839	QY	GCGGGGCCACCGACATCAATGGGTTACGCTCGGTCGGGCCAAACCTACGCGTCGAT	1898
1801	Db	GCGGGGCCACCGACATCAATGGGTTACGCTCGGTCGGGCCAAACCTACGCGTCGAT	1860
1899	QY	GTGGTCGGGTATGACCGCACCCAGGATGTCGGCTGCTGCAGCTGCGCGTGCCTGGTGC	1958
1861	Db	GTGGTCGGGTATGACCGCACCCAGGATGTCGGCTGCTGCAGCTGCGCGTGCCTGGTGC	1920
1959	QY	CTCCGTCGCGCGCGATCGGTGCGCGGTCGCGGTTGGTAGCCGCTGTCGCGATGGGC	2018
1921	Db	CTCCGTCGCGCGCGATCGGTGCGCGGTCGCGGTTGGTAGCCGCTGTCGCGATGGGC	1980
2019	QY	AACAGCGGTGGGACGGCGGAACGCCCGTCGGTGCCTCGCAGGTGGTTCGCGCTCGGC	2078
1981	Db	AACAGCGGTGGGACGGCGGAACGCCCGTCGGTGCCTCGCAGGTGGTTCGCGCTCGGC	2040
2079	QY	CAAAACCGTCGAGCGTCGGATTCCGTACCCGCTGCCGAAGACATGTAAACGGTTGATC	2138
2041	Db	CAAAACCGTCGAGCGTCGGATTCCGTACCCGCTGCCGAAGACATGTAAACGGTTGATC	2100
2139	QY	CAGTTGATGCCCGCATCAGCCCGGTGATTCGGGGGGGCCGCTCGTCAACGCCCTAGGA	2198
2101	Db	CAGTTGATGCCCGCATCAGCCCGGTGATTCGGGGGGGCCGCTCGTCAACGCCCTAGGA	2160
2199	QY	CAGGTGTCGGTATGAACACGGCGCGTCC	2228
2161	Db	CAGGTGTCGGTATGAACACGGCGCGTCC	2190

RESULT 8

```

US-10-369-983-6
; Sequence 6, Application US/10369983
; Publication No. US2003023593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-005081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2637
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-Erd14)

```

	Query Match	95.6%	Score 2186.8	DB 16	Length 2637
	Best Local Similarity	99.9%	Pred. No. 0		
	Matches 2188	Conservative 0	Mismatches 2	Indels 0	Gaps 0
Qy	39	CATATGATCACCATTCACCATCACCGCCGCGTCGGATAACTTCCAGCTGTCCACGGT	98		
Db	1	CATATGATCACCATTCACCATCACCGCCGCGTCGGATAACTTCCAGCTGTCCACGGT	60		

Qy	99	GGGACGGGATTGCGCATTTCCGATTCGGATCGGCGAGGCGATGGCGATCGCGGCGCAGATCCGATCG	158
Db	61	GGGACGGGATTGCGCATTTCCGATTCGGGCGAGGCGATGGCGATCGCGGCGCAGATCCGATCG	120
Qy	159	GGTGGGGGTTCACCCACCGTTTCATATCGGGCTACCGCCTTCTCGGCTTGGGTGTGTGC	218
Db	121	GGTGGGGGTTCACCCACCGTTTCATATCGGGCTACCGCCTTCTCGGCTTGGGTGTGTGC	180
Qy	219	GACAAACAACGGCAACGGCGCAGAGTCCAAACGGTGGTTCGGGAGCGCTTCGGCGGCAAGT	278
Db	181	GACAAACAACGGCAACGGCGCAGAGTCCAAACGGTGGTTCGGGAGCGCTTCGGCGGCAAGT	240
Qy	279	CTCGGCATCTCCACCGGCGAGTGCATACGGCGTCCAGCGGCTCCGATCAACTCGGCC	338
Db	241	CTCGGCATCTCCACCGGCGAGTGCATACGGCGTCCAGCGGCTCCGATCAACTCGGCC	300
Qy	339	ACCGCATGCGGACCGCGTTTAAACGGGCATATCCCGGTGAACGTCTCGGTGACCTGG	398
Db	301	ACCGCATGCGGACCGCGTTTAAACGGGCATATCCCGGTGAACGTCTCGGTGACCTGG	360
Qy	399	CAAAACAAGTCGGCGGCGACGGTACAGAGGACGTGACATTCGCGGAGGACCCCGGCC	458
Db	361	CAAAACAAGTCGGCGGCGACGGTACAGAGGACGTGACATTCGCGGAGGACCCCGGCC	420
Qy	459	GAATTATGTTGATTTTCGGGCGTTTACCACCGGAGATCAACTCCGCGAGGATGTACGCC	518
Db	421	GAATTATGTTGATTTTCGGGCGTTTACCACCGGAGATCAACTCCGCGAGGATGTACGCC	480
Qy	519	GGCCCGGGTTGGCCCTCGTGTGGCGCGGCTCAGATGTGGGACAGCTGCGGAGTGAC	578
Db	481	GGCCCGGGTTGGCCCTCGTGTGGCGCGGCTCAGATGTGGGACAGCTGCGGAGTGAC	540
Qy	579	CTGTTTTTCGSCCGCGTCGGCGTTTCAGTCCGTGGTCTGGGCTCTGACGCTGGGGTCTGTG	638
Db	541	CTGTTTTTCGSCCGCGTCGGCGTTTCAGTCCGTGGTCTGGGCTCTGACGCTGGGGTCTGTG	600
Qy	639	ATAGTTCTGTCGGCGTCTGATGTGTGGCGCGGCTCGCGTATGTGGCTGGATGACG	698
Db	601	ATAGTTCTGTCGGCGTCTGATGTGTGGCGCGGCTCGCGTATGTGGCTGGATGACG	660
Qy	699	GTACACGCGGGCAGCCGAGCTGACCGCGCCAGTCCGGTTCGTCGGCGGCTAC	758
Db	661	GTACACGCGGGCAGCCGAGCTGACCGCGCCAGTCCGGTTCGTCGGCGGCTAC	720
Qy	759	GAGACGGGTATGGGTGACGTTGCCCGCCCGCGGTGATCGCCGAGAACCGTGTGAACTG	818
Db	721	GAGACGGGTATGGGTGACGTTGCCCGCCCGCGGTGATCGCCGAGAACCGTGTGAACTG	780
Qy	819	ATGATTTGATAGGACCAACCTTTTGGGCGAAAACACCCCGCGATCGCGGTCAACGAG	878
Db	781	ATGATTTGATAGGACCAACCTTTTGGGCGAAAACACCCCGCGATCGCGGTCAACGAG	840
Qy	879	GCCGATATACCGCGAGATGTGGGCCCAAGACCGCCCGCGCATGTTTGGCTACGCGCGGGG	938
Db	841	GCCGATATACCGCGAGATGTGGGCCCAAGACCGCCCGCGCATGTTTGGCTACGCGCGGGG	900
Qy	939	ACGGCGACGCGGACGGGACGTTGTCTGCTTCGAGGAGCGCGCGGAGATGACACGGCG	998
Db	901	ACGGCGACGCGGACGGGACGTTGTCTGCTTCGAGGAGCGCGCGGAGATGACACGGCG	960
Qy	999	GGTGGGCTCTTCAGACAGGCGCGCGGTTCGAGGAGCGCTCCGACACCGCGCGGGGAAAC	1058
Db	961	GGTGGGCTCTTCAGACAGGCGCGCGGTTCGAGGAGCGCTCCGACACCGCGCGGGGAAAC	1020
Qy	1059	CAGTTGATGAACAATGTGCCCGCAGGCGCTGCAACAGCTGCGCCAGCCACGCAAGGACCC	1118
Db	1021	CAGTTGATGAACAATGTGCCCGCAGGCGCTGCAACAGCTGCGCCAGCCACGCAAGGACCC	1080
Qy	1119	ACGCTTCTTCCAGCTGGGTGCTGTGGAAGACGCTCTCGCCGATCGGTTCGCGCATC	1178
Db	1081	ACGCTTCTTCCAGCTGGGTGCTGTGGAAGACGCTCTCGCCGATCGGTTCGCGCATC	1140
Qy	1179	AGCAACATGTGTGATGGCGCAACAACCATGTTCGATGATCAACCTCGGGTGTGTGATG	1238

```
Db 1141 AGCAACATGTTGATGGCCAAACACACATGTCGATGACCAACTCGGTGTGTCGATG 1200
Qy 1239 ACCAACCTTTAGCTCGATGTTGAAGGCTTTGCTCCGGCGCGGCCGCCCAAGGCGGTG 1298
Db 1201 ACCAACCTTTAGCTCGATGTTGAAGGCTTTGCTCCGGCGCGGCCGCCCAAGGCGGTG 1260
Qy 1299 CAATCCGCGCGCAAAACGGGTCGCGGCGATGAGCTCGTGGGCACTCGCTGGGTTCT 1358
Db 1261 CAATCCGCGCGCAAAACGGGTCGCGGCGATGAGCTCGTGGGCACTCGCTGGGTTCT 1320
Qy 1359 TCGGCTCTGGCGGTGGGTGGCGCCCAACTTGGGTTCGGCGCGCTCGGTTCGTTG 1418
Db 1321 TCGGCTCTGGCGGTGGGTGGCGCCCAACTTGGGTTCGGCGCGCTCGGTTCGTTG 1380
Qy 1419 TCGGTGCGGCGAGCTGGGCGCGCGCAACACGAGCAGTCAACCGCGCGCGCGCGCGCTG 1478
Db 1381 TCGGTGCGGCGAGCTGGGCGCGCGCAACACGAGCAGTCAACCGCGCGCGCGCGCGCTG 1440
Qy 1479 CCGCTGACACGCTGACACGCGCGCGGAAAGAGGCGCGGAGATGCTGGGCGGCGCTG 1538
Db 1441 CCGCTGACACGCTGACACGCGCGCGGAAAGAGGCGCGGAGATGCTGGGCGGCGCTG 1500
Qy 1539 CCGGTGGGCGAGATGGGCGCGCGAGGCGGTGGGTGCTAGTGGTGTGCTGCTGTTCCG 1598
Db 1501 CCGGTGGGCGAGATGGGCGCGCGAGGCGGTGGGTGCTAGTGGTGTGCTGCTGTTCCG 1560
Qy 1599 CCGGACCTTATGATGCGCATCTCCGCGAGCGCGGATATCGGCCCGCGCGCTTG 1658
Db 1561 CCGGACCTTATGATGCGCATCTCCGCGAGCGCGGATATCGGCCCGCGCGCTTG 1620
Qy 1659 TCGCAGACCGGTTGCGCGATTTCCCGCGCTGCCCCGAGCTCGCGCGCTGCGGTCGCC 1718
Db 1621 TCGCAGACCGGTTGCGCGATTTCCCGCGCTGCCCCGAGCTCGCGCGCTGCGGTCGCC 1680
Qy 1719 CAGTGGGCGCACAGGTGTCACATCAACACCAACTGGGCTACACACGCGGTGGG 1778
Db 1681 CAGTGGGCGCACAGGTGTCACATCAACACCAACTGGGCTACACACGCGGTGGG 1740
Qy 1779 GCCGGGACCGGATCGTCATGATCCCAACGGTGTGCTGTGACCAACCAACGATGATC 1838
Db 1741 GCCGGGACCGGATCGTCATGATCCCAACGGTGTGCTGTGACCAACCAACGATGATC 1800
Qy 1839 GCGGCGCGCACCATCATGCTTACGCTGCGGTTCGCGGCGCAACCTACGCGTTCGAT 1898
Db 1801 GCGGCGCGCACCATCATGCTTACGCTGCGGTTCGCGGCGCAACCTACGCGTTCGAT 1860
Qy 1899 GTGTCGCGGTATGACCGCACCGGATGTCGCGGTGTCGAGCTGCGCGGTGCGCGTGGC 1958
Db 1861 GTGTCGCGGTATGACCGCACCGGATGTCGCGGTGTCGAGCTGCGCGGTGCGCGTGGC 1920
Qy 1959 CTCGCGTGGCGGCGATCGGTGGCGGCTCGCGGTGGTGGAGCCGCTGTCGCGATGGGC 2018
Db 1921 CTCGCGTGGCGGCGATCGGTGGCGGCTCGCGGTGGTGGAGCCGCTGTCGCGATGGGC 1980
Qy 2019 AACAGCGGTGGCGAGCGGAAACCGCGGTGCGGTGCTGCGAGGCTGGTTCGCGTTCGC 2078
Db 1981 AACAGCGGTGGCGAGCGGAAACCGCGGTGCGGTGCTGCGAGGCTGGTTCGCGTTCGC 2040
Qy 2079 CAATCCGTGAGGCGTTCGATTCGCTGACCGGTTCGCGAGAGACATTTGAACGGTTGATC 2138
Db 2041 CAATCCGTGAGGCGTTCGATTCGCTGACCGGTTCGCGAGAGACATTTGAACGGTTGATC 2100
Qy 2139 CAGTTCCATCCCGGATCCAGCGCGGTGATTCGCGGCGGCGCGCTGCTCAACGCGCTAGGA 2198
Db 2101 CAGTTCCATCCCGGATCCAGCGCGGTGATTCGCGGCGGCGCGCGCTGCTCAACGCGCTAGGA 2160
Qy 2199 CAGGTGTCGGTATGAACACGCGCGGTGCC 2228
Db 2161 CAGGTGTCGGTATGAACACGCGCGGTGCC 2190
```

```
US-10-098-732A-64
; Sequence 64, Application US/10098732A
; Publication NO. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 64
; LENGTH: 2808
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72F-MAPS
; OTHER INFORMATION: (795f) fusion construct, TB MTB72F (Ra12-FB9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-64
```

```
Query Match 95.6%; Score 2186.8; DB 15; Length 2808;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 39 CATATGATCATCAACATCAACATCACACGCGCGCTCCGATCACTTCCAGCTGTCCAGGGT 98
Db 1 CATATGATCATCAACATCAACATCACACGCGCGCTCCGATCACTTCCAGCTGTCCAGGGT 60
Qy 99 GGGCAGGATTCGCGATTCGATCGGCGAGCGATCGGATCGGCGGCGAGATCCGATCG 158
Db 61 GGGCAGGATTCGCGATTCGATCGGCGAGCGATCGGATCGGCGGCGAGATCCGATCG 120
Qy 159 GGTGGGGGTCAACCAACCGTTTCATATCGGCGCTACCGCTTCCCTCGGCTTGGGTGTGTC 218
Db 121 GGTGGGGGTCAACCAACCGTTTCATATCGGCGCTACCGCTTCCCTCGGCTTGGGTGTGTC 180
Qy 219 GACAAACAGCGACGCGGCGACGAGTCACGCGTGTGTCGAGCGCTCCGCGGCGAAGT 278
Db 181 GACAAACAGCGACGCGGCGACGAGTCACGCGTGTGTCGAGCGCTCCGCGGCGAAGT 240
Qy 279 CTCGCGATCTCCACCGCGAGCTGATCACCGCGGTTCAGCGCGCTCCGATCAACTCGGCC 338
Db 241 CTCGCGATCTCCACCGCGAGCTGATCACCGCGGTTCAGCGCGCTCCGATCAACTCGGCC 300
Qy 339 ACCGCGATTCGCGAGCGCTTACCGGCGATCACTCCGCTGACGTCATCTCGGTGACCTGG 398
Db 301 ACCGCGATTCGCGAGCGCTTACCGGCGATCACTCCGCTGACGTCATCTCGGTGACCTGG 360
Qy 399 CAATCCAGTTCGCGCGCGACGCGTACAGGAACTGACATTCGCGGAGGAGACCGCGGCC 458
Db 361 CAATCCAGTTCGCGCGCGACGCGTACAGGAACTGACATTCGCGGAGGAGACCGCGGCC 420
Qy 459 GAATTCATGTGTGATTTTCGGGGCGTTTACCGCGGAGATCAACTCCGCGAGGATGTACGCC 518
Db 421 GAATTCATGTGTGATTTTCGGGGCGTTTACCGCGGAGATCAACTCCGCGAGGATGTACGCC 480
Qy 519 GCGCGGGTTCGCGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCTGGCGAGTGAC 578
Db 481 GCGCGGGTTCGCGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCTGGCGAGTGAC 540
Qy 579 CTGTTTTTCGCGCGCGTTCGCGGTTTCAGTTCGCTGCTGGGGTCTGACGCTGGGTCTGG 638
Db 541 CTGTTTTTCGCGCGCGTTCGCGGTTTCAGTTCGCTGCTGGGGTCTGACGCTGGGTCTGG 600
Qy 639 ATAGGTTTCGCGCGCGTTCGATGTTGGTGGCGGCGCTCCGCGGTATGTGGGTGATGAGC 698
```

Db 601 ATAGGTTCTGGCGGCTCTGATGTGGCGCGGCTCGCCGTATGTGGCTGATGAGC 660
QY 699 GTACCGCGGAGCGAGTCGACCGCCGCCAGGTCGCGGTTCGTGCGGCGCCCTAC 758
Db 661 GTACCGCGGAGCGAGTCGACCGCCGCCAGGTCGCGGTTCGTGCGGCGCCCTAC 720
QY 759 GAGACGGCTATGGCTGACGTCGCGCCCGCGGTGATCGCCGAGAACCGTGTCAATG 818
Db 721 GAGACGGCTATGGCTGACGTCGCGCCCGCGGTGATCGCCGAGAACCGTGTCAATG 780
QY 819 ATGATTCTGATAGCGACCAACCTCTTTGGGGCAAAACACCCGCGGATCGCGTCAACGAG 878
Db 781 ATGATTCTGATAGCGACCAACCTCTTTGGGGCAAAACACCCGCGGATCGCGTCAACGAG 840
QY 879 GCGGATACGGCGAGATGTGGGCGGAGAGCGCGCGCGATGTTGGCTACGCCCGGGG 938
Db 841 GCGGATACGGCGAGATGTGGGCGGAGAGCGCGCGCGATGTTGGCTACGCCCGGGG 900
QY 939 AGGGCGACGGCGACGCGACGTTGCTGCGGTTTCGAGGAGCGCGGAGATGACCAAGCGCG 998
Db 901 AGGGCGACGGCGACGCGACGTTGCTGCGGTTTCGAGGAGCGCGGAGATGACCAAGCGCG 960
QY 999 GGTGGGCTCTCGAGCAGCGCGCGGCTCGAGAGGCTTCGACACCGCGCGCGCGAAC 1059
Db 961 GGTGGGCTCTCGAGCAGCGCGCGGCTCGAGAGGCTTCGACACCGCGCGCGCGAAC 1020
QY 1059 CAGTTGATCAACATGTGCGCCAGCGCTGCAACAGCTGCGCCAGCGCGAGGCGCAC 1118
Db 1021 CAGTTGATCAACATGTGCGCCAGCGCTGCAACAGCTGCGCCAGCGCGAGGCGCAC 1080
QY 1119 AGCGCTCTTTCAAGCTGGGTGGCTGTGGAGAGAGCTGCGCGCATCGTTCGCGCATC 1178
Db 1081 AGCGCTCTTTCAAGCTGGGTGGCTGTGGAGAGAGCTGCGCGCATCGTTCGCGCATC 1140
QY 1179 AGCAATCGTGTGATGTCGACCAACACACATGTCGATGACCACTCGCGGTGTCGATG 1238
Db 1141 AGCAATCGTGTGATGTCGACCAACACACATGTCGATGACCACTCGCGGTGTCGATG 1200
QY 1239 ACCAACACCTTGAGCTCGATGTTGAAGGCTTTGCTCCGCGCGCGCGCGCGCGCGTG 1298
Db 1201 ACCAACACCTTGAGCTCGATGTTGAAGGCTTTGCTCCGCGCGCGCGCGCGCGTG 1260
QY 1299 CAAACCGCGCGCAAAACCGGGTCCGGGCGATGAGCTCGCTGGGAGCTCGTGGGTCT 1358
Db 1261 CAAACCGCGCGCAAAACCGGGTCCGGGCGATGAGCTCGCTGGGAGCTCGTGGGTCT 1320
QY 1359 TCGGCTCGGCGGTGGGTGGCGCAACTTGGGTGGGCGCGCTCGGTCGCTGCTG 1418
Db 1321 TCGGCTCGGCGGTGGGTGGCGCAACTTGGGTGGGCGCGCTCGGTCGCTGCTG 1380
QY 1419 TCGGTGCGCGAGCTGGGCGCGGCGCAACCGAGGAGTCAACCGCGCGCGCGCGCTG 1478
Db 1381 TCGGTGCGCGAGCTGGGCGCGGCGCAACCGAGGAGTCAACCGCGCGCGCGCGCTG 1440
QY 1479 CGGCTACACGCTGACCGCGCGCGGAGAGGCGCGGAGATGCTGGCGGGGCTG 1538
Db 1441 CGGCTACACGCTGACCGCGCGCGGAGAGGCGCGGAGATGCTGGCGGGGCTG 1500
QY 1539 CCGGTGGGCGAGATGGGCGCGGAGGCGGCTGTTGGGCTCAGTGTGCTGCTGCTGCTG 1598
Db 1501 CCGGTGGGCGAGATGGGCGCGGAGGCGGCTGTTGGGCTCAGTGTGCTGCTGCTGCTG 1560
QY 1599 CCGGACCTATGATGCGGCTTCTCGGCGCGCGGAGATGCTGGCGCGCGGCTG 1658
Db 1561 CCGGACCTATGATGCGGCTTCTCGGCGCGCGGAGATGCTGGCGCGCGGCTG 1620
QY 1659 TCGCAGGCGCGGTTCCCGGCTTCCCGGCTGCGGCTCGACCGCTCGCGGATGTCGCG 1718
Db 1621 TCGCAGGCGCGGTTCCCGGCTTCCCGGCTGCGGCTCGACCGCTCGCGGATGTCGCG 1680
QY 1719 CAACTGGGCGCACAGTGTGTCACATCAACAACTGGGCTACAAACCGCGCGGCTG 1778

Db 1681 CAAGTGGGCGCACAGGTGGTCAACATCAACACCAACTGGGCTACAAACCGCGCTGGC 1740
QY 1779 GCGGGAACCGGCTATGTCATCGATCCCAACCGGTGTCGTGCTGACCAACACCGTATC 1838
Db 1741 GCGGGAACCGGCTATGTCATCGATCCCAACCGGTGTCGTGCTGACCAACACCGTATC 1800
QY 1839 GCGGGAACCGGCTATGTCATCGATCCCAACCGGTGTCGTGCTGACCAACCGTATC 1898
Db 1801 GCGGGAACCGGCTATGTCATCGATCCCAACCGGTGTCGTGCTGACCAACCGTATC 1860
QY 1899 GTGTCGGGTATGACCCACCGAGATGTCGGGTGTCGAGTTCGCGGTGTCGGGTGTC 1958
Db 1861 GTGTCGGGTATGACCCACCGAGATGTCGGGTGTCGAGTTCGCGGTGTCGGGTGTC 1920
QY 1959 CTGCGGTGCGCGGCGATCGGTGTCGGGTGTCGGGTGTCGGGTGTCGGGTGTCGGGT 2018
Db 1921 CTGCGGTGCGCGGCGATCGGTGTCGGGTGTCGGGTGTCGGGTGTCGGGTGTCGGGT 1980
QY 2019 AACAGCGGTGGGCGAGGCGGAAACCGCGGTGTCGGGTGTCGGGTGTCGGGTGTCGGGT 2078
Db 1981 AACAGCGGTGGGCGAGGCGGAAACCGCGGTGTCGGGTGTCGGGTGTCGGGTGTCGGGT 2040
QY 2079 CAACCGGTGCGGCGTGGATTCGCTGACCGGTGCGGAGAGACATTGAACGGGTTCATC 2138
Db 2041 CAACCGGTGCGGCGTGGATTCGCTGACCGGTGCGGAGAGACATTGAACGGGTTCATC 2100
QY 2139 CAGTTCGATGCGCGGATCCAGCGCGGTGATTTCGGGCGCGCGCGGTGCTGCTCAACGGCTAGGA 2198
Db 2101 CAGTTCGATGCGCGGATCCAGCGCGGTGATTTCGGGCGCGCGCGGTGCTGCTCAACGGCTAGGA 2160
QY 2199 CAGTTCGATGCGCGGATCCAGCGCGGTGATTTCGGGCGCGCGCGGTGCTGCTCAACGGCTAGGA 2228
Db 2161 CAGTTCGATGCGCGGATCCAGCGCGGTGATTTCGGGCGCGCGCGGTGCTGCTCAACGGCTAGGA 2190

RESULT 10

US-10-369-983-5

; Sequence 5, Application US/10369983

; Publication No. US20030235593A1

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis

; FILE REFERENCE: 014059-009081US

; CURRENT APPLICATION NUMBER: US/10/369,983

; PRIOR FILING DATE: 2003-02-18

; PRIOR APPLICATION NUMBER: US 60/357,351

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 2808

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: fusion protein

; OTHER INFORMATION: R95F (MTB72P-MAPS)

US-10-369-983-5

Query Match 95.6%; Score 2186.8; DB 16; Length 2808;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 39 CATATGATCATCAATCAACCATCAACCGGCGCGGTCCGATAACTTCCAGCTGTCCAGGGT 98

Db 1 CATATGATCATCAATCAACCATCAACCGGCGCGGTCCGATAACTTCCAGCTGTCCAGGGT 60

QY 99 GGGCAGGATTCGCATTCGATCGGCGAGCGATGCGATCGCGGCGAGATCCGATCCGATCG 158

Db 61 GGGCAGGATTCGCATTCGATCGGCGAGCGATGCGATCGCGGCGAGATCCGATCCGATCG 120

QY 159 GGTGGGGGTCAACCAACGGTTTATATCGGGGCTTACCGCCTTCTCGGCTTGGGTGTGTC 218
DB 121 GGTGGGGGTCAACCAACGGTTTATATCGGGGCTTACCGCCTTCTCGGCTTGGGTGTGTC 180
QY 219 GACAAACGCGCAACGGCGCAGAGTCCAGCGGTGGTGGGAGCGCTCGGGCGGCAAGT 278
DB 181 GACAAACGCGCAACGGCGCAGAGTCCAGCGGTGGTGGGAGCGCTCGGGCGGCAAGT 240
QY 279 CTCGGCATCTCACCGGCGAGTGTATCACCGCGGTGACCGCGCTCCGATCAACTCGGCC 338
DB 241 CTCGGCATCTCACCGGCGAGTGTATCACCGCGGTGACCGCGCTCCGATCAACTCGGCC 300
QY 339 ACCCGGATGGGAGCGGCTTAAACGGGATCATCCCGGTGACGTCTCTCGGTGACCTGG 398
DB 301 ACCCGGATGGGAGCGGCTTAAACGGGATCATCCCGGTGACGTCTCTCGGTGACCTGG 360
QY 399 CAACCAAGTCCGGCGGCAACCGGTACAGGGAAAGTGAATTTGGCCGAGGACCCCGGCC 458
DB 361 CAACCAAGTCCGGCGGCAACCGGTACAGGGAAAGTGAATTTGGCCGAGGACCCCGGCC 420
QY 459 GAATTCATGTGGATTTTCGGGGCTTACCAACGGAGATCAACTCCCGGAGGATGTAGGCC 518
DB 421 GAATTCATGTGGATTTTCGGGGCTTACCAACGGAGATCAACTCCCGGAGGATGTAGGCC 480
QY 519 GCGCGGGTTCGGGCTCGCTGGTGGCGCGGCTCAGATGTGGACAGCGTGGCGAGTGAC 578
DB 481 GCGCGGGTTCGGGCTCGCTGGTGGCGCGGCTCAGATGTGGACAGCGTGGCGAGTGAC 540
QY 579 CTGTTTTCGGCGGCTCGGCGTTTCAGTGGTGGTCTTGGGCTCTGACGGTGGGGTCTGG 638
DB 541 CTGTTTTCGGCGGCTCGGCGTTTCAGTGGTGGTCTTGGGCTCTGACGGTGGGGTCTGG 600
QY 639 ATAGGTTCTCGCGGCTCTGATGGTGGCGCGGCTCTCGGCTATGTGGGCTGGATGAC 698
DB 601 ATAGGTTCTCGCGGCTCTGATGGTGGCGCGGCTCTCGGCTATGTGGGCTGGATGAC 660
QY 699 GTACCGCGGCGAGGCGAGTGTACCGCGCGCGCGCGGCTTGTGCGCGCGGCGCTAC 758
DB 661 GTACCGCGGCGAGGCGAGTGTACCGCGCGCGCGCGGCTTGTGCGCGCGGCGCTAC 720
QY 759 GAGACGGGATAGGCTGAGCGTGGCGCGCGCGCGGCTGATCCCGAGAACCGTGTGACTG 818
DB 721 GAGACGGGATAGGCTGAGCGTGGCGCGCGCGCGGCTGATCCCGAGAACCGTGTGACTG 780
QY 819 ATGATTCTGATAGGCAACCACTCTTGGGGCAAAAACACCCCGCGCATCGCGGTCAACGAG 878
DB 781 ATGATTCTGATAGGCAACCACTCTTGGGGCAAAAACACCCCGCGCATCGCGGTCAACGAG 840
QY 879 GCCGAATACGGCGAGATGTGGGCGCAAGACCGCGCGCGATTTGGCTACGCGCGGCG 938
DB 841 GCCGAATACGGCGAGATGTGGGCGCAAGACCGCGCGCGATTTGGCTACGCGCGGCG 900
QY 939 ACGCGGACGCGAGCGGAGCTTGTCTCGGTTCGAGGAGCGCGGAGATGACACGCGG 998
DB 901 ACGCGGACGCGAGCGGAGCTTGTCTCGGTTCGAGGAGCGCGGAGATGACACGCGG 960
QY 999 GGTGGGTCTCTCGAGGAGCGCGCGGCTCGAGGAGCGCTTCGACACCGCGCGGCGAAC 1058
DB 961 GGTGGGTCTCTCGAGGAGCGCGCGGCTCGAGGAGCGCTTCGACACCGCGCGGCGAAC 1020
QY 1059 CAGTTGATGACATGTGCCAGGCGCTGACACAGCTGGCCAGCGCCAGCGAGGACCC 1118
DB 1021 CAGTTGATGACATGTGCCAGGCGCTGACACAGCTGGCCAGCGCCAGCGAGGACCC 1080
QY 1119 ACGCTTCTTCCAGCTGGGTGGCTGTGAAGACGCTCTCGCGCATCGGTGCGCGATC 1178
DB 1081 ACGCTTCTTCCAGCTGGGTGGCTGTGAAGACGCTCTCGCGCATCGGTGCGCGATC 1140
QY 1179 AGCAACATGTGTGATGGCGCAACCAACCATGTGATGACCAACTCGGGTGTGCTGATG 1238
DB 1141 AGCAACATGTGTGATGGCGCAACCAACCATGTGATGACCAACTCGGGTGTGCTGATG 1200
QY 1239 ACCAACACCTTGAGTCTGATGTTGAAGGGCTTGTGCTCGCGCGCGCGCGCCAGGCGCTG 1298

RESULT 11

US-10-369-983-11
; Sequence, 11, Application US/10369983
; Publication No. US2003023559A1
; GENERAL INFORMATION:

DB 1201 ACCAACACCTTGAGTCTGATGTTGAAGGGCTTGTCTCGGGCGGGCGCCAGCGCTG 1260
QY 1299 CAAACCGGGCGCAAAACGGGTCCGGCGCATGAGTCTGGGCGAGTCTCGTGGGTCT 1358
DB 1261 CAAACCGGGCGCAAAACGGGTCCGGCGCATGAGTCTGGGCGAGTCTCGTGGGTCT 1320
QY 1359 TCGGCTCTGGGCGGTGGGTGGCCCAACTTGGGTGGGCGGCTCGGTGGGTCTGTTG 1418
DB 1321 TCGGCTCTGGGCGGTGGGTGGCCCAACTTGGGTGGGCGGCTCGGTGGGTCTGTTG 1380
QY 1419 TCGGTCGCGCAGGCTGGGCGCGCCCAACAGAGCATCACCCGGCGGCGCGCGCTG 1478
DB 1381 TCGGTCGCGCAGGCTGGGCGCGCCCAACAGAGCATCACCCGGCGGCGCGCGCTG 1440
QY 1479 CGGCTGACAGCTGACAGCGCGCGGAAAGAGGGCCCGGCGAGATGCTGGGCGGCTG 1538
DB 1441 CGGCTGACAGCTGACAGCGCGCGGAAAGAGGGCCCGGCGAGATGCTGGGCGGCTG 1500
QY 1539 CGGTCGGGCGAGTGGGCGCGCGCGGTGGTGGGTCTAGTGGTCTGCTGGGTCTCG 1598
DB 1501 CGGTCGGGCGAGTGGGCGCGCGCGGTGGTGGGTCTAGTGGTCTGCTGGGTCTCG 1560
QY 1599 CGGCGACCTTATGTGATCGCGCATTTCTCCGCGAGCGCGCGATATCGCCCGCGCTTG 1658
DB 1561 CGGCGACCTTATGTGATCGCGCATTTCTCCGCGAGCGCGCGATATCGCCCGCGCTTG 1620
QY 1659 TCGCAGGACCGGTTCGCGCACTTCCCGCGCTGCCCTTCGACCGCTCGCGCATGTCGCC 1718
DB 1621 TCGCAGGACCGGTTCGCGCACTTCCCGCGCTGCCCTTCGACCGCTCGCGCATGTCGCC 1680
QY 1719 CAAAGTGGGCGCACAGGTGTCTCAACATCAACACCAAACTGGGCTACAAACACCGCGTGGC 1778
DB 1681 CAAAGTGGGCGCACAGGTGTCTCAACATCAACACCAAACTGGGCTACAAACACCGCGTGGC 1740
QY 1779 CCGCGGACCGCATCTGTCATCGATCCCAACCGGTGTGGTGTGACAAACACCGTGTATC 1838
DB 1741 GCGCGGACCGCATCTGTCATCGATCCCAACCGGTGTGGTGTGACAAACACCGTGTATC 1800
QY 1839 GCGGCGCACCGCATCAATCGTTCAGCTCCGCTCCGCGCAAACTACGGGCTCGAT 1898
DB 1801 GCGGCGCACCGCATCAATCGTTCAGCTCCGCTCCGCGCAAACTACGGGCTCGAT 1860
QY 1899 GTGGTCGGGTATGACCGCACCCAGGATGTCCGGTGTGTCAGCTGCGCGGTGCGCGTGGC 1958
DB 1861 GTGGTCGGGTATGACCGCACCCAGGATGTCCGGTGTGTCAGCTGCGCGGTGCGCGTGGC 1920
QY 1959 CTGCGCTCGGCGCGATCGGTGGGCGGTCCGGTGGTGGAGCCCGCTCGTCCGATGGC 2018
DB 1921 CTGCGCTCGGCGCGATCGGTGGGCGGTCCGGTGGTGGAGCCCGCTCGTCCGATGGC 1980
QY 2019 AACACGGTGGCGAGGCGGAAACCGCTCGGTGCTTGGCAGGGTGGTCCGCTCGGC 2078
DB 1981 AACACGGTGGCGAGGCGGAAACCGCTCGGTGCTTGGCAGGGTGGTCCGCTCGGC 2040
QY 2079 CAAACCGTGCAGGCGTGGATTCGCTGACCGGTGCCGAAGACATTTGAACGGGTGATC 2138
DB 2041 CAAACCGTGCAGGCGTGGATTCGCTGACCGGTGCCGAAGACATTTGAACGGGTGATC 2100
QY 2139 CAGTTGATGCGCGATCCAGCCCGGTGATTCGGGCGGCGCGCTCGTCAACGGGCTAGGA 2198
DB 2101 CAGTTGATGCGCGATCCAGCCCGGTGATTCGGGCGGCGCGCTCGTCAACGGGCTAGGA 2160
QY 2199 CAGTTGCTGGTATGAACACCGCGCTCC 2228
DB 2161 CAGTTGCTGGTATGAACACCGCGCTCC 2190

APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
CURRENT FILING DATE: 2003-02-18
PRIOR APPLICATION NUMBER: US 60/357,351
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 3060
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: fusion protein
OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-11

Query Match 95.6%; Score 2186.8; DB 16; Length 3060;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	39	CATATGCATCACCATCACCATCACACGGCGCGTCCGATAACTTCAGCTGTCACAGGT	98
DB	1	CATATGCATCACCATCACCATCACACGGCGCGTCCGATAACTTCAGCTGTCACAGGT	60
QY	99	GGGAGGGATTCGCCATTCCGATTCGGGCGAGGCGATGGCGATCGCGGCGAGATCCGATCG	158
DB	61	GGGAGGGATTCGCCATTCCGATTCGGGCGAGGCGATGGCGATCGCGGCGAGATCCGATCG	120
QY	159	GGTGGGGGTCAACCAAGTTCATATCGGGCTACCGCTTCTCGGCTGGGTGGTGTTC	218
DB	121	GGTGGGGGTCAACCAAGTTCATATCGGGCTTACCGCTTCTCGGCTGGGTGGTGTTC	180
QY	219	GACAAACAGCGCAACCGCGCACAGTCCAAACGGTGGTCCGAGCGCTCCGCGCAAGT	278
DB	181	GACAAACAGCGCAACCGCGCACAGTCCAAACGGTGGTCCGAGCGCTCCGCGCAAGT	240
QY	279	CTCGGATCTCACCGGCGAGTATCACCGGGTCAACGGGCTCCGATCAACTCGGCC	338
DB	241	CTCGGATCTCACCGGCGAGTATCACCGGGTCAACGGGCTCCGATCAACTCGGCC	300
QY	339	ACCGGATGGCGAGCGCTTAACGGGATCATCCCGGTGAAGTCACTCGGTGACCTGG	398
DB	301	ACCGGATGGCGAGCGCTTAACGGGATCATCCCGGTGAAGTCACTCGGTGACCTGG	360
QY	399	CRAACCAAGTCGGCGGCGAGGTACAGGAAAGTGAATGGCCGAGGACCCCGCGCC	458
DB	361	CRAACCAAGTCGGCGGCGAGGTACAGGAAAGTGAATGGCCGAGGACCCCGCGCC	420
QY	459	GAATTCATGGTATTCGGGGGTTTACACCGGATCAATCCCGGAGGATGACGCC	518
DB	421	GAATTCATGGTATTCGGGGGTTTACACCGGATCAATCCCGGAGGATGACGCC	480
QY	519	GGCCCGGGTTCGGCTCGTGGTGGCGGGGTTCAGATGGGACAGCGTGGCGAGTGAC	578
DB	481	GGCCCGGGTTCGGCTCGTGGTGGCGGGGTTCAGATGGGACAGCGTGGCGAGTGAC	540
QY	579	CTGTTTTCGGCGCGCTCGGGTTTCACTGGTGGTCTGGGGTTCAGCGTGGGTCTGG	638
DB	541	CTGTTTTCGGCGCGCTCGGGTTTCACTGGTGGTCTGGGGTTCAGCGTGGGTCTGG	600
QY	639	ATAGGTTTCGGCGGGTTCGATGGTGGCGGGCTCCCGGTATGTCGGTGGATGAGC	698
DB	601	ATAGGTTTCGGCGGGTTCGATGGTGGCGGGCTCCCGGTATGTCGGTGGATGAGC	660
QY	699	GTACCGCGGGGCGAGCGAGCTGAACCGCGCCAGGTCCGGGTCTCGCGCGGCTTAC	758
DB	661	GTACCGCGGGGCGAGCGAGCTGAACCGCGCCAGGTCCGGGTCTCGCGCGGCTTAC	720

QY	759	GAGACGGCTATGGCTGACGGTCCCGCGCGTGTGATCCCGAGAACCTGCTGAAC	818
DB	721	GAGACGGCTATGGCTGACGGTCCCGCGCGTGTGATCCCGAGAACCTGCTGAAC	780
QY	819	ATGATTTCTGATAGCGACCAACCTCTTGGGCAAAAACACCCCGCGATCGCGGTCAACG	878
DB	781	ATGATTTCTGATAGCGACCAACCTCTTGGGCAAAAACACCCCGCGATCGCGGTCAACG	840
QY	879	GCCGAATACGGCGAGATGTGGGCCCAAGACCGCGCGCGATGTTGGCTACGCGCGCG	938
DB	841	GCCGAATACGGCGAGATGTGGGCCCAAGACCGCGCGCGATGTTGGCTACGCGCGCG	900
QY	939	ACGGCGACGGCGACGGCGAGCTTCTCCCTTCGAGGAGCGCGCGAGATGACACAGCGCG	998
DB	901	ACGGCGACGGCGACGGCGAGCTTCTCCCTTCGAGGAGCGCGCGAGATGACACAGCGCG	960
QY	999	GGTGGGTCTCTCGAGAGCGCGCGCGTTCGAGAGGCTTCGACACCGCGCGCGCGAAC	1058
DB	961	GGTGGGTCTCTCGAGAGCGCGCGCGTTCGAGAGGCTTCGACACCGCGCGCGCGAAC	1020
QY	1059	CAGTTGATGAACATGTGCCCGAGCGCTCAACAGCTGGCCAGCCAGCCAGCGGACCC	1118
DB	1021	CAGTTGATGAACATGTGCCCGAGCGCTTCACAGCTGGCCAGCCAGCGGAGCCAGCC	1080
QY	1119	ACGCTTTCTTCAAGCTGGGTGGCTTGTGAAGACGGTCTTCGCGGATTCGGTCCCGATC	1178
DB	1081	ACGCTTTCTTCAAGCTGGGTGGCTTGTGAAGACGGTCTTCGCGGATTCGGTCCCGATC	1140
QY	1179	AGCAACATGTGTGATGGCGCAACCAACACATCTCGATGACCACTCGGGTGTTCGATG	1238
DB	1141	AGCAACATGTGTGATGGCGCAACCAACACATCTCGATGACCACTCGGGTGTTCGATG	1200
QY	1239	ACCAACACCTTGAAGCTGTGAGGGGCTTCTCCGGCGCGCGCGCGCGCGCGCGCG	1298
DB	1201	ACCAACACCTTGAAGCTGTGAGGGGCTTCTCCGGCGCGCGCGCGCGCGCGCGCG	1260
QY	1299	CAAAACCGCGCGCAAAACCGGGTCCGGCGGATGAGCTTCGCTGGGCGAGCTCGCTGGT	1358
DB	1261	CAAAACCGCGCGCAAAACCGGGTCCGGCGGATGAGCTTCGCTGGGCGAGCTCGCTGGT	1320
QY	1359	TCGGTCTGGGCGGTGGGTCGCGCAACTTGGGTTCGGCGCGGCTCGGTTCGGTTCG	1418
DB	1321	TCGGTCTGGGCGGTGGGTCGCGCAACTTGGGTTCGGCGCGGCTCGGTTCGGTTCG	1380
QY	1419	TCGGTTCGGCGAGGCTGGGCGCGCGCAACAGGACGTCACCCCGCGCGCGCGCGCG	1478
DB	1381	TCGGTTCGGCGAGGCTGGGCGCGCGCAACAGGACGTCACCCCGCGCGCGCGCGCG	1440
QY	1479	CCGCTGACCGAGCTGACAGCGCGCGGAGAGGCGCGCGCGAGATGCTGGGCGGGCTG	1538
DB	1441	CCGCTGACCGAGCTGACAGCGCGCGGAGAGGCGCGCGAGATGCTGGGCGGGCTG	1500
QY	1539	CCGCTGGGCGAGATGGGCGCGCGCGGCTGGTGGGTTCAGTGTGTCTGCTGTTCCG	1598
DB	1501	CCGCTGGGCGAGATGGGCGCGCGCGGCTGGTGGGTTCAGTGTGTCTGCTGTTCCG	1560
QY	1599	CCGCGACCTTATGTGATGCGCATTCGCGGACCGCGGATTCGCGCGCGCGCGCTTG	1658
DB	1561	CCGCGACCTTATGTGATGCGCATTCGCGGACCGCGGATTCGCGCGCGCGCGCTTG	1620
QY	1659	TCGAGAGACCGGTTTCGCGACTTCCCGCGCTGCCCTTCGACCCGCTCGCGATGCTGC	1718
DB	1621	TCGAGAGACCGGTTTCGCGACTTCCCGCGCTGCCCTTCGACCCGCTCGCGATGCTGC	1680
QY	1719	CAAGTGGGCGCAAGGTGGTCAACATCAACACCAAAATGGGCTACAACAACCGCGTGGC	1778
DB	1681	CAAGTGGGCGCAAGGTGGTCAACATCAACACCAAAATGGGCTACAACAACCGCGTGGC	1740
QY	1779	GCCGGGACCGGATCGATCCCAACCGTCTCGTGTGACCAACCAACACCGTGTATC	1838
DB	1741	GCCGGGACCGGATCGATCCCAACCGTCTCGTGTGACCAACCAACACCGTGTATC	1800
QY	1839	CCGGGCGCGACCGACATCAATGCTTTCAGCGTTCGGCTCCGCGCGCAAACTACGGCTCG	1898

342 GCGATGGCGGACGCGCTTAAACGGGCGATCATCCCGGTGACGTCACTCGGTGACCTGGCAA 401
Db
301 GCGATGGCGGACGCGCTTAAACGGGCGATCATCCCGGTGACGTCACTCGGTGACCTGGCAA 360
Qy
402 ACCAAGTCGGGCGGACCGGTCACAGGAAACGTGACATTTGGCCGAGGAGCCCGCGCGAA 461
Db
361 ACCAAGTCGGGCGGACCGGTCACAGGAAACGTGACATTTGGCCGAGGAGCCCGCGCGAA 420
Qy
462 TTCATGGTGGATTTTCGGGCGGTTTACACCGGAGATCAACTCCCGCGAGGATGTACCGCGGC 521
Db
421 TTCATGGTGGATTTTCGGGCGGTTTACACCGGAGATCAACTCCCGCGAGGATGTACCGCGGC 480
Qy
522 CCGGTTTCGGGCGGTCGTTGGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTACCGCTG 581
Db
481 CCGGTTTCGGGCGGTCGTTGGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTACCGCTG 540
Qy
582 TTTTCGGGCGGTCGCGCTTTTCAGTCGGTGGCTTGGGCTCAGCGTGGGCTGGGATA 641
Db
541 TTTTCGGGCGGTCGCGCTTTTCAGTCGGTGGCTTGGGCTCAGCGTGGGCTGGGATA 600
Qy
642 GGTTCGTCGGCGGCTTCATGTGGCGGCGGCTCGCGGTATGTGGCGTGGAGTACGCTG 701
Db
601 GGTTCGTCGGCGGCTTCATGTGGCGGCGGCTCGCGGTATGTGGCGTGGAGTACGCTG 660
Qy
702 ACCGCGGCGGCGGCTGACCGCGCGGCGGCTCGCGGTTCGCGGCGGCTTACGAG 761
Db
661 ACCGCGGCGGCGGCTGACCGCGCGGCGGCTCGCGGTTCGCGGCGGCTTACGAG 720
Qy
762 ACCGCGTATGGGCTGACCGGTCGCGCGGCGGCTCGCGGAGAACCGTGTGAACTGATG 821
Db
721 ACCGCGTATGGGCTGACCGGTCGCGCGGCGGCTCGCGGAGAACCGTGTGAACTGATG 780
Qy
822 ATTCTGATAGGACCAACCTTTTGGGCGAAACACCCCGCGGATCGCGGTCAACGAGGCC 881
Db
781 ATTCTGATAGGACCAACCTTTTGGGCGAAACACCCCGCGGATCGCGGTCAACGAGGCC 840
Qy
882 GAATACGGCGAGATGTGGCCCAAGACGCGCGCGATGTTGGCTACCGCGCGCGAGC 941
Db
841 GAATACGGCGAGATGTGGCCCAAGACGCGCGCGATGTTGGCTACCGCGCGCGAGC 900
Qy
942 GCGACGGCGACGCGGCTGCTGCTGCGGTCGAGAGGCGCGGAGATGACACGCGCGGT 1001
Db
901 GCGACGGCGACGCGGCTGCTGCTGCGGTCGAGAGGCGCGGAGATGACACGCGCGGT 960
Qy
1002 GGGCTCTCGAGCAGCGCGCGGCTCGAGGAGGCTCGACACCGCGCGCGGAGCACAG 1061
Db
961 GGGCTCTCGAGCAGCGCGCGGCTCGAGGAGGCTCGACACCGCGCGCGGAGCACAG 1020
Qy
1062 TTGATGAACAATCTGCCCCAGGCGCTGCAACAGCTGGCCCCAGCCCCAGCGGCGACCAAG 1121
Db
1021 TTGATGAACAATCTGCCCCAGGCGCTGCAACAGCTGGCCCCAGCCCCAGCGGCGACCAAG 1080
Qy
1122 CTTTCTTCAAGCTGGGTCGCTGTGAGAGACGCGTTCGCGGATCGGTTCGCGGATCAGC 1181
Db
1081 CTTTCTTCAAGCTGGGTCGCTGTGAGAGACGCGTTCGCGGATCGGTTCGCGGATCAGC 1140
Qy
1182 AACATGTTGCGATGCGCAACCAACCATGCTCGATGACCAACTCGCGGTGTGCGATGACC 1241
Db
1141 AACATGTTGCGATGCGCAACCAACCATGCTCGATGACCAACTCGCGGTGTGCGATGACC 1200
Qy
1242 AACACCTTTAGCTCGATGTTGAAGGCTTTGCTCCCGCGCGCGGCGCGCGCGGCGTGA 1301
Db
1201 AACACCTTTAGCTCGATGTTGAAGGCTTTGCTCCCGCGCGCGGCGCGCGCGGCGTGA 1260
Qy
1302 ACCGCGCGCAAAACGGGTCGCGGCGATGAGCTCGCTGGGAGCTCGCTGGGTTCTTCG 1361
Db
1261 ACCGCGCGCAAAACGGGTCGCGGCGATGAGCTCGCTGGGAGCTCGCTGGGTTCTTCG 1320
Qy
1362 GGTTCGGCGGTCGGGTCGCGCGCAACTTGGGTCGGGCGGCTCGGTTCGTTGTCG 1421
Db
1321 GGTTCGGCGGTCGGGTCGCGCGCAACTTGGGTCGGGCGGCTCGGTTCGTTGTCG 1380
Qy
1422 GTGCGCGGCGCTTGGGCGCGCGCAACCGAGCTACCCCGCGGCGGCGGCTGCGC 1481

1381 GTCCGCGAGGCTTGGGCGCGGCAACACGAGCTACCCCCGCGCGCGGCGCTGCGC 1440
Qy
1482 CTGACGAGCTTACGAGCGCGGAGAAAGAGGCGCGGCGAGATGCTGGGCGGCTGCGC 1541
Db
1441 CTGACGAGCTTACGAGCGCGCGGAGAAAGAGGCGCGGCGAGATGCTGGGCGGCTGCGC 1500
Qy
1542 GTGGGCGAGATGGGCGCGAGGCGCGGTGGGCTCAGTGGTGTGCTGCGGTTCGCGCG 1601
Db
1501 GTGGGCGAGATGGGCGCGAGGCGCGGTGGGCTCAGTGGTGTGCTGCGGTTCGCGCG 1560
Qy
1602 CGACCTTATGTATGCGGCAATTCGCGGAGCGCGGATATCGCCCCGCGCGCTTGTG 1661
Db
1561 CGACCTTATGTATGCGGCAATTCGCGGAGCGCGGATATCGCCCCGCGCGCTTGTG 1620
Qy
1662 CAGACCGGTTCCCGCACTTCCCGCGCTCGACCCGCTCGACCCGCTCGCGATGGTGGCCAA 1721
Db
1621 CAGACCGGTTCCCGCACTTCCCGCGCTCGACCCGCTCGACCCGCTCGCGATGGTGGCCAA 1680
Qy
1722 GTGGGCGAGATGGGCTCAACATCAACCAAACTGGGCTACAAACGCGCGGTGGGCGCC 1781
Db
1681 GTGGGCGAGATGGGCTCAACATCAACCAAACTGGGCTACAAACGCGCGGTGGGCGCC 1740
Qy
1782 GGGACCGGCTATCGTATCGATCCCAACGCTGTGCTGTGACCAACACGATGATGCG 1841
Db
1741 GGGACCGGCTATCGTATCGATCCCAACGCTGTGCTGTGACCAACACGATGATGCG 1800
Qy
1842 GGGCGCACGACATCAATGCTTCAGGCTCGGCTCGGCGCAACCTACGCGCTCGATG 1901
Db
1801 GGGCGCACGACATCAATGCTTCAGGCTCGGCTCGGCGCAACCTACGCGCTCGATG 1860
Qy
1902 GTCCGCTATGACCCACACCGAGATGTCGCGGTGTGACGCTGCGCGGTGCGGCTG 1961
Db
1861 GTCCGCTATGACCCACACCGAGATGTCGCGGTGTGACGCTGCGCGGTGCGGCTG 1920
Qy
1962 CCGTCGCGGCGATCGGTGGCGGCTCGCGGTGGTGGAGCCGCTCGGTGGGCGAAC 2021
Db
1921 CCGTCGCGGCGATCGGTGGCGGCTCGCGGTGGTGGAGCCGCTCGGTGGGCGAAC 1980
Qy
2022 AGCGTGGCGAGGCGGAAACGCGCGCTGCGTGGCGAGGCTGCGCGCTCGGCGAA 2081
Db
1981 AGCGTGGCGAGGCGGAAACGCGCGCTGCGTGGCGAGGCTGCGCGCTCGGCGAA 2040
Qy
2082 ACCGTGAGGCGCTCGGATTCGCTGACCGGTGCGGAGAGACATTTGAACGCGTTGATCAG 2141
Db
2041 ACCGTGAGGCGCTCGGATTCGCTGACCGGTGCGGAGAGACATTTGAACGCGTTGATCAG 2100
Qy
2142 TTGATGCGCGGATCCAGCCGCTGATTCGGCGGCGCGCTCGTCAACGCGCTAGGACAG 2201
Db
2101 TTGATGCGCGGATCCAGCCGCTGATTCGGCGGCGCGCTCGTCAACGCGCTAGGACAG 2160
Qy
2202 GTGGTGGTATGAACACGCGCGCTGCTAG 2231
Db
2161 GTGGTGGTATGAACACGCGCGCTGCTAG 2190

RESULT 15
US-10-098-732A-17
; Sequence 17, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Gudarian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80

[illegible]

Qy	2022	ACGGTGGCAGGCGGAACGCCGCTGGGTGCTTGGCAGGTTGGTCGGCTGGCCAA	2081
Db	1981	ACGGTGGCAGGCGGAACGCCGCTGGGTGCTTGGCAGGTTGGTCGGCTGGCCAA	2040
Qy	2082	ACCGTGCAGGCGTCCGATTTCGCTGACCGTGCCGAAGAGACATTGAACGGTTGATCCAG	2141
Db	2041	ACCGTGCAGGCGTCCGATTTCGCTGACCGTGCCGAAGAGACATTGAACGGTTGATCCAG	2100
Qy	2142	TTTCGATGCCCGCATCCAGCCCGGTGATTCCGGCGGGCCCGTCTGTCAACGGCTAGGACAG	2201
Db	2101	TTTCGATGCCCGCATCCAGCCCGGTGATTCCGGCGGGCCCGTCTGTCAACGGCTAGGACAG	2160
Qy	2202	GTGGTCGGTATGAACACGCGCCCGTCTCTAG	2231
Db	2161	GTGGTCGGTATGAACACGCGCCCGTCTCTAG	2190

Search completed: July 4, 2004, 00:29:18
Job time : 1017 secs

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes the need for transparency and accountability in financial reporting.

2. The second part of the document outlines the various methods and techniques used to collect and analyze data. It includes a detailed description of the experimental procedures and the statistical analysis performed.

3. The third part of the document presents the results of the study. It includes a series of tables and graphs that illustrate the findings of the research. The data shows a clear trend in the relationship between the variables studied.

4. The fourth part of the document discusses the implications of the findings and provides recommendations for future research. It suggests that further studies should be conducted to explore the underlying mechanisms of the observed phenomena.

5. The fifth part of the document is a conclusion that summarizes the main points of the study. It reiterates the importance of the research and the need for continued investigation in this field.

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2004, 17:20:36 ; Search time 5931 Seconds
(without alignments)

11712.366 Million cell updates/sec

Title: US-09-597-796C-11

Perfect score: 2287

Sequence: 1 tctagaataattttgtttta.....ggngtaacaaagcccgaaa 2287

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_esti:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.2	3.4	1628	29	CG757066 P052-2-A0
c 2	76.4	3.3	1406	29	CG756569 P051-4-B0
3	76.2	3.3	625	14	CD937289 OV.106112
c 4	76.2	3.3	935	29	CNS006XK

5	75	3.3	611	13	BU102503
6	75	3.3	611	13	CA064610
7	74.6	3.3	925	29	CNS0091P
c 8	73.6	3.2	1516	12	EGS0984
9	72.8	3.2	935	29	CNS006XK
10	72.6	3.2	949	29	AG171092
c 11	72.4	3.2	925	29	CNS0091P
c 12	72	3.1	1798	29	AG171124
c 13	71.8	3.1	1956	29	CG754548
c 14	71.2	3.1	880	29	CG458121
c 15	70.8	3.1	1160	12	EGS44853
c 16	70	3.1	550	14	CG869131
c 17	70	3.1	897	29	AG046101
c 18	69.6	3.0	1061	13	EX391246
19	69.4	3.0	932	29	CNS0072Q
20	69.4	3.0	1116	28	BZ569478
21	68.8	3.0	1203	29	CNS015Y4
c 22	68.6	3.0	1079	29	AG080398
23	68.4	3.0	862	29	AG036151
c 24	68.2	3.0	1218	13	BQ876717
c 25	68	3.0	1151	14	CK210749
c 26	67.8	3.0	289	29	U82114
27	67.8	3.0	1057	13	EX349688
c 28	67.6	3.0	750	29	AG149322
30	67.4	2.9	982	13	EX415111
c 31	67.2	2.9	807	14	CF149720
32	67.2	2.9	821	29	AG063084
33	67.2	2.9	1474	28	BZ569821
34	67	2.9	641	13	CA109290
c 35	66.8	2.9	932	29	CNS0072Q
c 36	66.6	2.9	924	13	BU187694
c 37	66.6	2.9	1121	29	AG062320
c 38	66.6	2.9	1798	29	AG171124
c 39	66.2	2.9	868	29	AG061755
c 40	66	2.9	742	29	CC688263
c 41	66	2.9	851	14	CB908312
42	66	2.9	978	29	CC735401
43	66	2.9	1075	29	AG073983
c 44	66	2.9	1200	13	EX456467
c 45	66	2.9	1288	13	BQ678719

ALIGNMENTS

RESULT 1
CG757066
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

CG757066 1628 bp DNA linear GSS 24-OCT-2003
P052-2-A02.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
Genomic survey sequence.
CG757066
CG757066.1 GI:37985257
GSS.
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 1628)
Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode Pristionchus
pacificus
Mol. Genet. Genomics 269 (5), 715-722 (2003)
22835951
12884007
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de

FEATURES		Class: BAC ends.		Location/Qualifiers	
source	1..1628	Best Local Similarity	40.8%;	Pred. No. 0.23;	Length 1628;
	/organism="Pristionchus pacificus"	Matches 642;	Conservative 0;	Mismatches 923;	Indels 9;
	/mol_type="genomic DNA"				Gaps 4;
	/strain="California"				
	/db_xref="taxon:54126"				
	/clone_lib="Ppa EcoRI BAC library"				
	/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."				
ORIGIN		Query Match		Score 77.2;	
					DB 29;
					Length 1628;
QY	519	GGCCCGGTTTCGGCTCCTCGTGGCGCGCTCAGATGTGGACACGCTGGCGAGTGAC	578		
DB	45	GGGNGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG	104		
QY	579	CTGTTTTCGCGCGCTCGCGCTTTTCAGTCGTGTCTGGGTCTGACGTTGGGTCTGTG	638		
DB	105	AGGNGGNGGGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG	164		
QY	639	ATAGTTTCGTCGGCGGTTCTGATGTTGGCGCGCTCGCGCTGATGTGGGTGATGAGC	698		
DB	165	GNGGNGGGGGGCG	224		
QY	699	GTACCGGGGGAGCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	758		
DB	225	GGGNGGCGCGCGGGGCG	284		
QY	759	GAGACGGGTATGCGCTGACGTTGCCCGCGCGCTGATCGCGAGAACCGTGTGAAC	818		
DB	285	CGGGGGGCG	344		
QY	819	ATGATTCTGATAGCGACAACTTTTGGGGCAAAACACCCCGCGATTCGCGGTCAA	878		
DB	345	CG	404		
QY	879	GCCGAATACGCGAGATGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	938		
DB	405	CGCGGGGCG	464		
QY	939	ACGCGCAGCGCGACGCGACGTTGCTGCTCGTTGAGGAGCGCGCGGAGATGAC	998		
DB	465	GGGCGCGGGGGGCG	524		
QY	999	GGTGGGCTCTCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1058		
DB	525	GGCGGGGCG	579		
QY	1059	CAGTTGATGAACATGTGCTCCCGCGCGCTGCAACAGCTGGCCCGCGCGCGCGCGCG	1118		
DB	580	GNGGG	639		
QY	1119	ACGCTTCTTCCAGCTGGTGGCTGTGGAAGACGGTCTGCG-CGATCGGTTCGCCAT	1177		
DB	640	GCGCGGGGCG	699		
QY	1178	CAGCAACATGTTGTCATGATGCGCAACACATGTGATGACCAACTCGGGTGTGTCAT	1237		
DB	700	GG	759		
QY	1238	GACCAACACTTGTAGCTCGATGTTGAAGGCTTTGTTCGCGCGCGCGCGCGCGCGCG	1297		
DB	760	GGGCGCGCGGG	819		
QY	1298	GCAACCGCGCGCAAAACCGGGTTCGGGGGATGAGTCTGCTG--GGCAGCTCGTGGT	1355		
DB	820	GNCGGGGCGCGGGCG	879		

QY	1356	TCTTCGGGTCTCGGGCTGGGTGGGTGGCGCCCAACTTGGGTGGCGCGCGCTCGGTCTCG	1415		
DB	880	GCG	939		
QY	1416	TTGTTCGGTCTCGCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG	1475		
DB	940	GGGGGCGCGGG	999		
QY	1476	CTGCGCTCTACCGCTGACAGCGCGCGCGGAAAGAGGGCGCGCGCGATGCTGGCGGG	1535		
DB	1000	GGCGGG	1059		
QY	1536	CTGCGCGTGGGCGACATGGGCGCCAGGCGCGTGGTGGGCTCAGTGGTGTGCTGCTGTT	1595		
DB	1060	GGGGCGGG	1119		
QY	1596	CGCGCGCGACCTATGTATGTCGCGCATTTCTCGGCGAGCGCGCGATATCGCCCCCGCG	1655		
DB	1120	NGCGGG	1179		
QY	1656	TTGTTCGCGAGGCGGTTTCGCGACTTTCGCGGCTTCCCGGCTGCCCCCTGACCCGTC	1715		
DB	1180	GCG	1239		
QY	1716	GCCCAAGTGGGCGCACAGGTGTCAACATCAACACCAAACTGGGCTACAACAACGCGCTG	1775		
DB	1240	NCCCGCGGG	1299		
QY	1776	GGCGCGGGGACCGGATCGTATCGATCCCAACGGTGTGCTGTGACCAACAACACGTTG	1835		
DB	1300	GGCG	1359		
QY	1836	ATCGCGCGCGCGCACACACATCAATGCTTCAGCTCGGCTCCCGCCAAACCTACGGGCTC	1895		
DB	1360	GGCGGG	1419		
QY	1896	GATGTGTGCGGATATGACCGCACCGAGATTCGCGGTGTGCTGCGGTGCGCGGTTC	1955		
DB	1420	GCG	1479		
QY	1956	GCGCTGCGGTTCGGGGCGGATCGGTG-CGCGCGTTCGCGTGTGTCAGCCCGCTCGCGAT	2014		
DB	1480	GCCCG	1539		
QY	2015	GGGCAACAGCGTGGGCGAGGCGGAAACCCCGCTGCGTGTGCTGCGAGGGTGTGTCGCT	2074		
DB	1540	GGCGGG	1599		
QY	2075	GGGCAACACCGTGC 2088			
DB	1600	CGCGGGGGCGCGCG 1613			
RESULT 2					
CG756569/C					
LOCUS	CG756569 1406 bp DNA linear GSS 24-OCT-2003				
DEFINITION	P051-4-B06.zb Ppa EcoRI BAC library Pristionchus pacificus genomic,				
	genomic survey sequence.				
ACCESSION	CG756569				
VERSION	CG756569.1 GI:37984265				
KEYWORDS	GSS.				
SOURCE	Pristionchus pacificus				
ORGANISM	Pristionchus pacificus				
	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;				
	Neodiplogasteridae; Pristionchus.				
REFERENCE	1 (bases 1 to 1406)				
AUTHORS	Srinivasan, J., Sinz, W., Jesse, T., Wiggers-Pereboite, L., Jansen, K.,				
	Buntjer, J., van der Meulen, M. and Sommer, R.J.				
TITLE	An integrated physical and genetic map of the nematode Pristionchus				
	pacificus				
JOURNAL	Mol. Genet. Genomics 269 (5), 715-722 (2003)				
MEDLINE	22835951				
PUBMED	12884007				

Db 83 CAGGAGCGCGCCGGTGCAGGTGGTCTCGCGGAGCGCGGGAAGCGGTGCGCGGTG 142
 QY 558 TGGACACAGCGTGGCGAGTGAACCTGTTTTCGCGCGCGTGGCGGTTCAGTTCGGTCTGG 617
 Db 143 GTGGCTTGGAGGTGGTTCAGCGCGAGGAGGAGTCTGGTGGTGGTGGCGAGG---AG 199
 QY 618 GGTCTGACGGTGGGTGCTGATAGTTCGTGCGCGGGTCTGATGCTGGCGCGGCTCG 677
 Db 200 GTTCAGGTGGTGGCTCGAGGTGGTCTGGTTCGCGTGGTGGTTCAGTTCGAGGTGGCGCG 259
 QY 678 CCGTATGTTGGGTGATGAGCGTCAACCGCGGGGAGCGCGAGTGAACCGCGCGCGGCTC 737
 Db 260 GAGAGGTGGCGGTTCAGCGGTGGTCTGGTGGTGGCGGAGGAGGTTCAGGCGGTG 319
 QY 738 CGGTTGCTGCGCGCGGTTCAGAGACGGCGGTATGGGCTGACGGTGGCGCGCGGTGATC 797
 Db 320 GCCTTGGAGGAGGTGTCGGATCAGGTGGTGGCTGGTGGTGGCGGAGGAGGTTCAG 379
 QY 798 GCCGAGAACCTGCTGACTGATGATTCGTATAGCAGCAACCTCTTGGGGCAAAACACC 857
 Db 380 GCGTGGCTTGGAGGAGGTGCGGTTCAGCGCGTGGCGGTGGTGGTGGGAGGAGGAG 439
 QY 858 CCGCGCATCGCGGTCAACGAGCGCGGATACGGCGGAGATGTGGGCCCAAGACGCGCGCG 917
 Db 440 GCTCGGTGGTGGTCTGGAGGCGGTTCAGGTGGAGGTGGCGGTTCAGGCGGTGGTTCG 499
 QY 918 ATGTTTGGCTACGCGCGGGGAGCGCGACCGCGGAGCGCGAGTTCGTGCGGTTCGAGGAG 977
 Db 500 GTGAGGAGCTCANGCGGTGGGTGTTGAGCGCGCGCGGAGGAGTGTGGACTCGCGCGGTG 559
 QY 978 GCGCGCGGAGTACGACGCGGTGGGCTC 1007
 Db 560 GAGCGCGGAGGCGGAGGCGCGGAGGCC 589

RESULT 4
 CNS006XK/c 935 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
 DEFINITION BACR14N09 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL066051
 VERSION AL066051.1 GI:4945019
 SOURCE GSS.
 ORGANISM Drosophila melanogaster (fruit fly)
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 935)
 AUTHORS Direct Submission
 TITLE Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoss in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 FEATURES Location/Qualifiers
 source 1..935

ORIGIN
 Query Match 3.3%; Score 76.2; DB 29; Length 935;
 Best Local Similarity 28.8%; Pred. No. 0.29;
 Matches 122; Conservative 114; Mismatches 187; Indels 0; Gaps 0;
 QY 1421 GTGTCGCGAGCGTGGCGCGGCGCAACGAGGAGTCAACCGCGCGGCGGCGGTGCCC 1480
 Db 927 GSGSSSSGSGCGSGSGGSCCCSCGCGCCSCSCSSSSSCSSSSSCSSCGCS 868
 QY 1481 GGTGACCAAGCTGACGCGCGGCGGAAAGAGGCGCGGCGAGATGCTGGCGCGGTGCC 1540
 Db 867 GCCSCGSGSSCCSCGCGSGSSGCGCGCGCGCGCGSCCSCSSCCGCGCGS 808
 QY 1541 GTGTCGCGAGATGGCGCGCGCGCGCGGTGGGCTCAGTGTGCTGCTGCTGCTGCTG 1600
 Db 807 CSCGCGCGCGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGSS 748
 QY 1601 GCGACCTTATGATGCGCGCATTTCTCGGCGAGCGCGGATATCGCCCGCGCGCTTGTG 1660
 Db 747 GSGCGSGSGSGSCSGCGSGSGSGCGCGCGCGCGCGCGCGCGCGCGCGCGS 688
 QY 1661 CAGGACCGGTTCGCGCGCTTCCCGCGCGTCCCTCGACCGCGCGCGATGTCGCGCCA 1720
 Db 687 GSCG 628
 QY 1721 AGTGGCGCGAGGTGTCAACATCAACAACTGGGCTACAAACCGCGCGCGCGCG 1780
 Db 627 GCG 569
 QY 1781 CCGGACCGCGCATCGTATCGATCCCAACGGTGTGCTGCTGACCAACACCGCTGATCGC 1840
 Db 567 GCGMCGAGSGKMSAGSGSGCGCGCGCGSGSGSGSGSGSGSGSGSGSGSGSGSG 508
 QY 1841 GGG 1843
 Db 507 GBK 505
 RESULT 5
 BUI02503
 LOCUS SCCAD1001A12.g Saccharum officinarum mRNA (Nogueira,F.T.S)
 DEFINITION Saccharum officinarum cDNA, mRNA sequence.
 ACCESSION BUI02503
 VERSION BUI02503.1 GI:32813826
 KEYWORDS EST.
 SOURCE Saccharum officinarum
 ORGANISM Saccharum officinarum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
 REFERENCE 1 (bases 1 to 611)
 AUTHORS Nogueira,F.T.S., de Rosa,V.E. Jr., Menossi,M., Ulian,E.C. and Aruda,P.
 TITLE RNA expression profiles and data mining of sugarcane response to low temperature
 JOURNAL Plant Physiol. 132 (4), 1811-1824 (2003)
 MEDLINE 22795309
 PUBMED 12913139
 COMMENT Contact: Nogueira FTS
 Bioinformatics Lab
 Organization for Nucleotide Sequencing and Analysis
 C.P. 6176, Campinas, SP 13083-970, Brazil
 Tel: 55 19 37881101
 Fax: 55 19 37881089
 Email: tebaldi@unicamp.br.

FEATURES

source
 1. .611
 Location/Qualifiers
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone_lib="Saccharum officinarum mRNA (Nogueira, F.T.S)"

ORIGIN

Query Match 3.3%; Score 75; DB 13; Length 611;
 Best Local Similarity 47.8%; Pred. No. 0.39;
 Matches 253; Conservative 0; Mismatches 270; Indels 6; Gaps 1;
 QY 566 CTGCGGAGTGAACCTGTTTCGGCGCGCTGCGCGCTCTGAT-----GGTGGCGGCGGCTCGCC 679
 Db 9 CATGGCGCGCGCTTGGAGGTGGCCCTCGCGGTGAGGTGTGAGCGGAGGCGGCGCTTGGT 68
 QY 626 GTTGGGCTGTGATAGCGTCAACCGCGGGCAGCCGAGTGTACCGCGCGCCAGGTCGG 739
 Db 69 GCAGCGGAGGAGCGGCGCTCGCGGTGAGGTGTGAGGTGTGAGCGGAGGCGGCTTGGT 128
 QY 740 GCTTCTGCGCGCGCTACGAGCGCGTATGGGTGAGCGTGCCTCGCGCGCGCTGATCGC 799
 Db 189 GCCGTGGAGGAGTGGACTAGCGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 248
 QY 800 CGAGAACCTGTGTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 859
 Db 249 CGAGGTACTGCGCGCTTGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 308
 QY 860 GCGGATCGCGTCAACGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 919
 Db 309 GCGGCGCGGAGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 428
 QY 980 GCGGATCGCGTCAACGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1039
 Db 429 TCGGAGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 488
 QY 1040 CGACACCGCGCGGCAACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1088
 Db 489 GGTGTTGGCGCGCGCAAGGAGGCGGCTTTCGGCGGAGGCGGCGGTGCGGTGCGGTGCGGTGCGGT 537

RESULT 6

CA064610
 LOCUS
 DEFINITION
 5', mRNA sequence.

ACCESSION

CA064610
 VERSION
 CA064610.1 GI:34916134

KEYWORDS

SOURCE

ORGANISM

Saccharum officinarum
 Saccharum officinarum

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 611)
 Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.
 The libraries that made SUCEST
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
 Contact: Arruda P
 Centro de Biologia Molecular e Engenharia Genetica
 Universidade Estadual de Campinas
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil
 Tel: 55 19 3788 1137
 Fax: 55 19 3788 1089
 Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found
 through the Brazilian Clone Collection Center (BCCC) at
 http://www.bcccenter.fcav.unesp.br

Plate: 001 row: A column: 12

Seq primer: T7 Promoter Primer.

Location/Qualifiers

1. .611

/organism="Saccharum officinarum"

/mol_type="mRNA"

/db_xref="taxon:4547"

/clone="SCCAD1001A12"

/lab_host="DH10B"

/clone_lib="AD1"

/notes="Organ: seedlings inoculated with Gluconacetobacter
 diazotrophicans; Vector: pSport1; Site: 1; SalI; Site: 2;
 Note: An unidirectional cDNA library generated from
 seedlings inoculated with Gluconacetobacter
 diazotrophicans. cDNA was prepared from polyA+ mRNA using
 Superscript Plasmid System Kit (Invitrogen). The
 double-strand cDNAs were fractionated in a sepharose
 CL-2B 40cm-columns and fragments sizing between 0.8 and
 1.5 Kb were directionally cloned into the vector. Details
 of each source of RNA and library construction can be
 obtained at http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 3.3%; Score 75; DB 13; Length 611;
 Best Local Similarity 47.8%; Pred. No. 0.39;
 Matches 253; Conservative 0; Mismatches 270; Indels 6; Gaps 1;
 QY 566 CGTGGCGAGTGAACCTGTTTCGGCGCGCTGCGCGTTCAGTCTGCTGCTGCGGTCTTGAC 625
 Db 9 CATGGCGCGCGCTTGGAGGTGGCCCTGGAGGTGGCGCGCGGTGCTCTCGTGGTGGT 68
 QY 626 GGTGGGCTGTGATAGTTGTCGGCGGTCTGAT-----GGTGGCGGCGGCTCGCC 679
 Db 69 GCAGCGGAGGAGCGGCGCTCGCGGTGAGGTGTGAGGTGTGAGCGGAGGCGGCGCTTGGT 128
 QY 680 GTATGTCGCGTGAATGAGCGTCAACCGCGGCGGCGGCGGCTGACGCGCGCCAGGTCGG 739
 Db 129 GGTGTCGCGCGGTGCTTGGAGGTGTGAGGTGTGAGGTGTGAGGTGTGAGGTGTGAGGTGTGAGGT 188
 QY 740 GGTGTCGCGCGGCGCTACGAGCGGCGGTATGGGCTGACGCTGCGCGCGCGCGGTGATCGC 799
 Db 189 GCGGTCGAGGAGGTGAGTACGCGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 248
 QY 800 CGAGAACCTGTGTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 859
 Db 249 GGAGTACTGCGCGTGCCTTGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 308
 QY 860 GGCGATCGCGTCAACGAGCGCGAATACGCGCGAGATGTGGGCGCGCAAGACGCGCGCAT 919
 Db 309 GGAGGAGGAGGCTTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 368
 QY 920 GTTTGGCTACG 979
 Db 369 GCGCGCGCGGAGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 428
 QY 980 GCGCGAGATGACACGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1039
 Db 429 TTCGAGGTGCG 488
 QY 1040 CGACACCG 1088
 Db 489 GGTGTTGGCGCGCGCAAGGAGGCGGCTTTCGGCGGAGGCGGCGGTGCGGTGCGGTGCGGTGCGGT 537

RESULT 7

CNS0091P

LOCUS

DEFINITION

925 bp DNA linear GSS 03-JUN-1999
 Drosophila melanogaster genome survey sequence TET3 end of BAC #
 BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.


```
QY 1760 CTACAAACACCGCGTGGGCGCGGACCGGAT--CGTCATCGATCCCAACGGTGTGCTG 1817
Db 785 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726
QY 1818 CTGACCAACACCGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1877
Db 725 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666
QY 1878 GGCACAACTACCGCGTGTGATGTCGGGTATGACC--GCACCCAGGATGTCGGGTGTC 1935
Db 665 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606
QY 1936 TGCAGTGTGCGGTGCGCGTGTGCGTGTGCGTGTGCGTGTGCGTGTGCGTGTGCGTGTG 1995
Db 605 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546
QY 1996 GTGAGCGCGTGTGCGGATGCGCAACAGCGGTGCGGAGCGCGGAAACGCGCGTGTGCGGTGTC 2055
Db 545 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 487
QY 2056 CTGGAGGCGTGTGCGCGTGTGCGCAACCGTGTGCGGATGTCGTCGTCGTCGTCGTCG 2115
Db 486 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 427
QY 2116 AAGAGACATTTGAACCGGTTGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 2175
Db 426 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 367
QY 2176 GCGCGTGTCTCAACGCGCTAGACAGTGTGTCG 2208
Db 366 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 334
```

```
RESULT 9
CNS006XK
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence 17 end of BAC #
BAC14N09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL066051
VERSION
AL066051.1 GI:4945019
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 935)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library RPCI-43
Unpublished
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE [E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr]
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Peter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and SST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
1..935
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
```

```
/clone="BAC14N09"
/clone_lib="RPCI-98"
/notes="end : T7"

ORIGIN

Query Match 3.2%; Score 72.8; DB 29; Length 935;
Best Local Similarity 35.6%; Pred. No. 0.88;
Matches 116; Conservative 78; Mismatches 130; Indels 2; Gaps 1;

QY 1275 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1334
Db 611 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 670
QY 1335 TCGGTGCGCGAGCTCGCTGCGGTTCCTTCGGGTCTCGGGCGCGTGGGTGGCGCGCAACTTGGGT 1394
Db 671 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728
QY 1395 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1454
Db 729 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788
QY 1455 GTACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1514
Db 789 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 848
QY 1515 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1574
Db 849 SGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 908
QY 1575 CTAGTGTGTGCTGCGGTGTTTCGCC 1600
Db 909 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 934
```

```
RESULT 10
AG171092
LOCUS
DEFINITION
Pan troglodytes DNA, clone: RP43-040E09.TU, genomic survey
sequence.
ACCESSION
AG171092
VERSION
AG171092.1 GI:16700770
KEYWORDS
GSS.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE
1
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library RPCI-43
Unpublished
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimbases@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library RPCI-43. This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TU
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI
Location/Qualifiers
1..949
/organism="Pan troglodytes"
/mol_type="genomic DNA"
```



```

RESULT 13
CG754548      1956 bp      DNA      linear      GSS 24-OCT-2003
LOCUS        P050-1-A11.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
DEFINITION   genomic survey sequence.
ACCESSION    CG754548
VERSION      CG754548.1 GI:37980151
KEYWORDS     GSS.
SOURCE       Pristionchus pacificus
ORGANISM     Pristionchus pacificus
              Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
              Neodiplogasteridae; Pristionchus.
REFERENCE    1 (bases 1 to 1956)
AUTHORS      Srinivasan, J., Sinz, W., Jesse, T., Wiggers-Perebolte, L., Jansen, K.,
              Buntjer, J., van der Meulen, M. and Sommer, R.J.
TITLE        An integrated physical and genetic map of the nematode Pristionchus
              pacificus
JOURNAL      Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE      22835951
PUBMED       12884007
COMMENT      Contact: Sommer RJ
              Evolutionary Biology
              Max-Planck-Institute for Developmental Biology
              Spemannstr. 37-39, Tuebingen D-72076, Germany
              Tel: 00497071601371
              Fax: 00497071601498
              Email: ralf.sommer@uebingen.mpg.de
              Class: BAC ends
              Location/Qualifiers
              1..1956
              /organism="Pristionchus pacificus"
              /mol_type="genomic DNA"
              /strain="California"
              /db_xref="taxon:54126"
              /clone_lib="Ppa EcoRI BAC Library"
              /note="The library was generated by a partial digest of
              the genomic DNA with EcoRI and cloning into the BAC
              vector."

FEATURES             source
Query Match          3.1%; Score 71.8; DB 29; Length 1956;
Best Local Similarity 43.3%; Pred. No. 1.4;
Matches 457; Conservative 0; Mismatches 583; Indels 16; Gaps 5;

QY 516 GCACGCGCGGCTTGGCTCTGCTGGTGGCGCGGCTCAGATGTGGACACGCTGGCGAGT 575
DB 880 GNNNGNNNGGGNNNNNNNGGGGGGGGGGGGNGGNGGNGGGGGCGCCCGCGCGGNCN 939
QY 576 GACCTGTTTCGCGCGCTCGGGGTTTCACTCGGTGCTCGGGTCTGACGGTGGGGTCG 635
DB 940 GNNNGGNNNGGGGGGGGGGNGCGCGCGGGGGGGGNGCGCGCGCGCGCGCGCGCC 999
QY 636 TGGATAGTTCGTCGCGCGGCTCTGATGGTGGCGCGGCTTCGCCGATGTGGCGTGGATG 695
DB 1000 GGCGCNCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1056
QY 696 AGCTACCGCGGGGAGCGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755
DB 1057 ---GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1113
QY 756 TACGAGACGGGATGATGGGCTGAAGGTGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 815
DB 1114 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1173
QY 816 CTGATGATTTGATAGGACCAACCTCTTGGGGCAAAACACCGCGCGGATCGCGGTCAAC 875
DB 1174 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1231
QY 876 GAGCGCGAATACGCGGAGATGTGGGCGCAAGACGCGCGCGCGCGCGCGCGCGCGCG 935
DB 1232 GCGNCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1291

```

```

QY 936 GCGACGGCGACGGCGACCGCGAC-----GTTGCTGCGGTTTCGAGAGGCGCGCGGAGATGA 990
DB 1292 GCNCGNCGGCGNCCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1351
QY 991 CAGCGCGGGTGGGTCTCTCAGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1050
DB 1352 CCGGGGGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1411
QY 1051 CGGCGAACCACTTGATGAACAATGTCGCCAGGCGCTGCAACAGCTGCGCCCGACGCCACGC 1110
DB 1412 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1471
QY 1111 AGGCGCACACCCCTTCTTCCAAAGCTGGGTGGCTGTGGAAGACGGTCTCGCGCGATCGGT 1170
DB 1472 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1531
QY 1171 CGCGCATCAGCAACATGGTGTGATGGCCACACACATGTGATGACCACTCGGGTG 1230
DB 1532 CCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1591
QY 1231 TGTGATGACCAACACCTTTGAGCTCGATGTTTGAAGGCTTTTGTCTCGCGCGCGCGCGCG 1290
DB 1592 CCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1651
QY 1291 AGGCGGTGCAAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1350
DB 1652 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1709
QY 1351 TGGGTTCTTTCGGGTCTGGGGCGTGGGGTGCGCGCAACTTGGGTGCGCGCGCGCGCGCG 1410
DB 1710 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1768
QY 1411 GTTCGTTGTCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1470
DB 1769 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1828
QY 1471 GGGCGCTCGCGCTGACGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1530
DB 1829 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1888
QY 1531 CGCGGCTGCGCGTGGGGCAGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1566
DB 1889 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1924

RESULT 14
CG458121/c
LOCUS        PUFXV40TBC ZM 0.6.1.0 KB Zea mays genomic clone ZMMB7a075H08,
DEFINITION   genomic survey sequence.
ACCESSION    CG458121
VERSION      CG458121.1 GI:34843121
KEYWORDS     GSS.
SOURCE       Zea mays
ORGANISM     Zea mays
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
              clade; Panicoideae; Andropogoneae; Zea.
              1 (bases 1 to 880)
REFERENCE    Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
              Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
              Bennetzen, J.
              Maize Genomics Consortium
              Unpublished (2003)
              Other GSSs: PUFXV40TDC
              Contact: Cathy Whitelaw
              tigr
              9712 Medical Center Drive, Rockville, MD 20850, USA
              Tel: 301-838-5843
              Fax: 301-838-0208
              Email: whitelaw@tigr.org
              Seq primer: TR
              Class: sheared ends.

```

FEATURES		Location/Qualifiers	
source			
	1. .880		
	/organism="Zea mays"		
	/mol_type="genomic DNA"		
	/strain="B73"		
	/db_xref="taxon:4577"		
	/clone_lib="ZM_0.6_1.0_KB"		
	/notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high		
	Cot selected genomic DNA library"		
ORIGIN			
Query Match	3.1%;	Score 71.2;	DB 29; Length 880;
Best Local Similarity	45.9%;	Pred. No. 1.5;	
Matches 244;	Conservative 0;	Mismatches 288;	Indels 0; Gaps 0;
QY	536	GCTGCTGGCCGCGCTCAGATGTGGGACACGCTGGCGAGTACCTGTTTTCGCCCGCTC	595
DB	682	GTTTGTGGCGGGAGGATATGCGGAGGCGGTGGCGGTGCGGTGACCTTGGCTTGGT	623
QY	596	GGCGTTTCAGTCGGTGTCTGGGCTCTGACGCTGGGCTGCTGATAGGTTCTGCGCGG	655
DB	622	GGTGGCCCTTGGGCACCGGTGAGGGTTTGAATTTGGGATTGAGAGGGAATGTTGGGGT	653
QY	656	TCTGATGGTGGCGCGGCTTCGCGTATGTGGCTGTGATGAGGCTCACCGCGGGCAGGC	715
DB	562	GCAGGTGGTGGAGGCGGCTCGGTGGTGGAGCTGGAGGTGCTTAGCGGAGCGGAGGG	503
QY	716	CGAGCTACCGCGCCAGTCCGGTTGCTGCGCGCGCTACGAGACGCGCTATGGCT	775
DB	502	CTTGTGGCGGTGGAGAGGTGGTTATTGCGGAGGTCTGCGGAGGCGCGCGCTTGGG	443
QY	776	GACGTCGCCCGCGGCTGATCGCGGAGAACCGTGTCTGAATCTGATAGCGAC	835
DB	442	GFTGTTCCGGTGGAGTCTCGCGGTGTTGAAGCGGAGGCGGGCTTGGTGGT	383
QY	836	CAACCTCTTGGGGCAAAACCCCGCGATGCGGTTCAACGAGCCGAAATCGCGCGAGAT	895
DB	382	GCGGTTGAGCGCGGTGCTTGTGGCGGAGGAGAGCTGTGTCGGCGAGCAGGTGCGGGT	323
QY	896	GTGGGCCAAGACGCCCGCGGATGTTGGCTACGCCGCGCGAGCGGCGGCGAGCGG	955
DB	322	GCGCGCTCTGGCGGAGCGCGGTGCGAGAGTGGCGCTGAGCGCGCTATGCGCGAGGT	263
QY	956	GACGTTCTGCGGTTTCGAGGAGGCGCGGAGATACCGCGCGGTGGGCTCTTCGAGCA	1015
DB	262	GCGGTGACAGAGCGCGCTATGTTGGAGGAGCTGGTGTGATACGGTGGAGGTGACGA	203
QY	1016	GCGCGCGCGGTTCAGAGGCTTCGACACCGCGCGCGGACAGTTGATG	1067
DB	202	GGTGGCGCGGAGCGGCTACCGCGGAGAGCTGGTGTGTCGCGGTGGT	151
RESULT 15			
BG844853/c			
LOCUS	BG844853 1160 bp mRNA linear EST 29-MAY-2001		
DEFINITION	1024008A08.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II		
KEYWORDS	Chlamydomonas reinhardtii cDNA, mRNA sequence.		
ACCESSION	BG844853		
VERSION	BG844853.1 GI:14226037		
KEYWORDS	EST.		
SOURCE	Chlamydomonas reinhardtii		
ORGANISM	Chlamydomonas reinhardtii		
REFERENCE	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;		
AUTHORS	1 (bases 1 to 1160)		
TITLE	Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,		
JOURNAL	McDermott, J.P., Silflow, C., Stern, D. and Surzycki, R.		
COMMENT	Analyses of the Chlamydomonas reinhardtii Genome: A Model,		
	Unicellular System for Analyzing Gene Function and Regulation in		
	Vascular Plants; project phase 2		
	Unpublished (2000)		
	Contact: Charles Hauser		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 27.0437 Seconds
(without alignments)
2747.774 Million cell updates/sec

Title: US-09-597-796c-26

Perfect score: 1306

Sequence: 1 VAWMSVTAGQALTAQVRV.....YHRDGGKYAXSGRRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1304	99.8	263	2 AAW32379	Mycobacte
2	1304	99.8	263	2 AAW32447	Mycobacte
3	1304	99.8	263	2 AAW64317	Mycobacte
4	1304	99.8	263	2 AAW61680	M. tuberc
5	1304	99.8	263	2 AAY32062	Mycobacte
6	1304	99.8	263	2 AAY38984	M. tuberc
7	1304	99.8	263	2 AAY39121	M. tuberc
8	1304	99.8	263	5 AAE29706	Mycobacte
9	1304	99.8	263	5 AAE17570	Mycobacte
10	1304	99.8	358	5 AAU74591	Antigenic
11	1187	90.9	391	2 AAW32381	Mycobacte
12	1187	90.9	391	2 AAW32449	Mycobacte
13	1187	90.9	391	2 AAW64335	Mycobacte
14	1187	90.9	391	2 AAW81702	M. tuberc
15	1187	90.9	391	2 AAY04778	Mycobacte
16	1187	90.9	391	2 AAY38989	M. tuberc
17	1187	90.9	391	2 AAY39132	M. tuberc
18	1187	90.9	391	4 AAU01888	M. tuberc
19	1187	90.9	391	5 AAE29707	Mycobacte
20	1187	90.9	391	5 AAE17571	Mycobacte
21	1187	90.9	596	2 AAY32070	Mycobacte
22	1187	90.9	596	5 AAE29710	Mycobacte
23	1187	90.9	596	5 AAE17574	Mycobacte
24	1187	90.9	599	5 AAU74599	Antigenic
25	1187	90.9	600	2 AAY32068	Mycobacte

26	1187	90.9	600	5 AAU74597	Antigenic
27	1187	90.9	723	7 ADA26354	Mycobacte
28	1187	90.9	723	4 AAO22142	Ral2-H9-3
29	1187	90.9	723	5 AAE29709	Mycobacte
30	1187	90.9	723	5 AAE17573	Mycobacte
31	1187	90.9	723	7 ADA26374	Mycobacte
32	1187	90.9	744	4 AAU01902	M. tuberc
33	1187	90.9	788	4 AAU01903	M. tuberc
34	1187	90.9	813	7 ADA26367	Mycobacte
35	1187	90.9	815	4 AAU01904	M. tuberc
36	1187	90.9	825	7 ADA26365	Mycobacte
37	1187	90.9	875	7 ADA26365	Mycobacte
38	1187	90.9	930	5 AAE29731	Mycobacte
39	1187	90.9	930	7 ADA26364	Mycobacte
40	1187	90.9	1010	7 ADA26356	Mycobacte
41	1187	90.9	1016	7 ADA26370	M. bovis
42	1187	90.9	1022	7 ADA26369	Mycobacte
43	1187	90.9	1154	7 ADA26368	Mycobacte
44	1182	90.5	394	2 AAY04779	Mycobacte
45	1182	90.5	729	5 AAE29708	Mycobacte

ALIGNMENTS

RESULT 1

AAW32379
ID AAW32379 standard; protein; 263 AA.
XX
AC AAW32379;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbH-9.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 254
FT /note= "Any amino acid"
XX
FN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
XX
PR 22-SEP-1995; 95US-0032136.
XX
PR 22-MAR-1996; 96US-00620280.
XX
PR 05-JUN-1996; 96US-00658800.
XX
PR 12-JUL-1996; 96US-00680573.
XX
PR (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX
PI Vedwick TH, Twardzik DR;
XX
DR WPI; 1997-192904/17.
XX
DR N-PSDB; AAT91432.
XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
XX
PT useful for diagnosis of M. tuberculosis infection.
XX
PS Example 3; Page 138-139; 190pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
XX
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
XX
CC variant differing only in conservative substitutions and/or
XX
CC modifications). The present sequence represents a M.tuberculosis antigen,

PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 XX Vedvick TS, Twardzik DR, Lodes MJ;
 DR WPI; 1998-251292/22.
 XX N-PSDB; AAV44371.
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 XX
 XX Example 3; Page 125-126; 250pp; English.
 PS
 XX This polypeptide comprises Mycobacterium tuberculosis antigen TbH-9. A
 CC DNA sequence (see AAV44371) coding for antigen TbH-9 was isolated from a
 CC M. tuberculosis strain H37Rv expression library using sera from patients
 CC having pulmonary or pleural tuberculosis. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAV64291-W64379) comprising an antigenic portion of a
 CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
 CC tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic kits for
 CC detecting M. tuberculosis infection in a patient using these
 CC polypeptides, antibodies or oligonucleotide probes and primers, for the
 CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS
 CC field)
 XX Sequence 263 AA;
 SQ

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 60
 DB 1 VAWMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPENTTSAGLLLEQAAAVEEASD 120
 DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPENTTSAGLLLEQAAAVEEASD 120
 QY 121 TAAANQLMNNVPOALKQALQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMTN 180
 DB 121 TAAANQLMNNVPOALKQALQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMTN 180
 QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSIGSSLGSGGGVAAANLGRAA 240
 DB 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSIGSSLGSGGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 4
 AAW81680
 ID AAW81680 standard; protein; 263 AA.
 AC AAW81680;
 XX
 XX 27-JAN-1999 (first entry)
 DT
 XX M. tuberculosis immunogenic polypeptide TbH-9.
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 XX vaccine; pharmaceutical; infection; diagnosis.
 XX Mycobacterium tuberculosis.
 OS
 XX Key Location/Qualifiers
 XX Misc-difference 254
 FT /label= unknown
 FT
 XX

PN W09816646-A2.
 XX 23-APR-1998.
 XX 07-OCT-1997; 97WO-US018293.
 XX 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00816112.
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64479.
 XX
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 XX Example 3b; Page 119-120; 230pp; English.
 PS This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis
 XX Sequence 263 AA;
 SQ

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 60
 DB 1 VAWMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPENTTSAGLLLEQAAAVEEASD 120
 DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPENTTSAGLLLEQAAAVEEASD 120
 QY 121 TAAANQLMNNVPOALKQALQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMTN 180
 DB 121 TAAANQLMNNVPOALKQALQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMTN 180
 QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSIGSSLGSGGGVAAANLGRAA 240
 DB 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSIGSSLGSGGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 5
 AAY32062
 ID AAY32062 standard; protein; 263 AA.
 XX
 XX AAY32062;
 AC
 XX 17-JAN-2000 (first entry)
 DT
 XX Mycobacterium tuberculosis antigen TbH9.
 XX Tuberculosis; antigen; fusion protein; TbH9; diagnosis; therapy; vaccine;
 XX immunogen.
 XX Mycobacterium tuberculosis.
 OS
 XX Key Location/Qualifiers
 XX

FT Misc-difference 254
 XX /note= "not identified"

PN WO9951748-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US007717.
 XX
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky YAW, Alderson M, Campos-Neto A;
 XX
 DR WPI; 1999-601610/51.
 XX

XX New fusion proteins useful for diagnosis, prevention and treatment of
 FT tuberculosis.
 XX

PS Claim 1; Fig 4C-D; 83pp; English.
 XX

CC This sequence represents the Mycobacterium tuberculosis antigen TbH9. The
 CC invention provides fusion proteins (see AAY32059-71) containing at least
 CC 2 M. tuberculosis antigens such as TbH9, e.g. Mb32A (see AAY32059) and a
 CC TbH9-Tb38-1 fusion. The new fusion proteins are useful as vaccines for
 CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
 CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),
 CC monitoring of disease progression, and treatment of tuberculosis. They
 CC are more effective immunogens than mixtures of the individual protein
 CC components
 XX

SQ Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 1 VAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAAVEASD 120
 DB 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAAVEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPI SNMVSMMANNHSMTN 180
 DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPI SNMVSMMANNHSMTN 180
 QY 181 SGVSNMTLSSMLKGFAPAAAQAVQTAQNGVRAVMSLSSGLSSGLGGVAAANLGRAA 240
 DB 181 SGVSNMTLSSMLKGFAPAAAQAVQTAQNGVRAVMSLSSGLSSGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 6

AAY38984
 ID AAY38984 standard; protein; 263 AA.

XX
 AC AAY38984;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen protein TbH-9.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX

OS Mycobacterium tuberculosis.

XX
 PN WO9942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US003265.
 XX
 PR 18-FEB-1998; 98US-00024753.
 PR 05-MAY-1998; 98US-00072596.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AAZ19069.

XX New polypeptide comprising antigenic portions of M. tuberculosis.

PS Example 3; Page 160-161; 323pp; English.
 XX

CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX

SQ Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 1 VAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAAVEASD 120
 DB 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAAVEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPI SNMVSMMANNHSMTN 180
 DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPI SNMVSMMANNHSMTN 180
 QY 181 SGVSNMTLSSMLKGFAPAAAQAVQTAQNGVRAVMSLSSGLSSGLGGVAAANLGRAA 240
 DB 181 SGVSNMTLSSMLKGFAPAAAQAVQTAQNGVRAVMSLSSGLSSGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 7

AAY39121
 ID AAY39121 standard; protein; 263 AA.

XX
 AC AAY39121;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis antigen TbH-9 amino acid sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX

OS Mycobacterium tuberculosis.

[illegible]

AC AAEL1570;
 XX 22-APR-2002 (first entry)
 DE Mycobacterium species MTB39 (TbH9) protein #1.
 KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB39; TbH9 protein.
 XX Mycobacterium sp.
 OS
 XX
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 254 /label= Unknown
 FT /note= "Encoded by NAG"
 FN WO200198460-A2.
 XX
 XX
 PD 27-DEC-2001.
 XX
 XX 20-JUN-2001; 2001WO-US019959.
 XX
 XX 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0285737P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX
 PI Skeiky Y, Reed S, Alderson M;
 XX WPI; 2002-147798/19.
 DR N-PSDB; AAD28340.
 DR
 XX
 FT Composition comprising MTB39 antigen and MTB32A antigen from
 FT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 83; Page 100; 136pp; English.
 XX
 CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are also used as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB39 (TbH9) protein
 XX
 SQ Sequence 263 AA;
 Query Match 99.8%; Score 1304; DB 5; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105; Indels 0; Gaps 0;
 Matches 263; Conservative 0; Mismatches 0;
 QY 1 VAMSVTAGQAEJTAQVRVAAAYATAYGLTVPPVIAENRAELMILIAATLLGQNTPA 60
 Db 1 VAMSVTAGQAEJTAQVRVAAAYATAYGLTVPPVIAENRAELMILIAATLLGQNTPA 60
 QY 61 IANVEAYGEMWQAQDAAMFGYAAATATATATLLPPEAPENTSGAGLLEQAAVVEEASD 120
 Db 61 IANVEAYGEMWQAQDAAMFGYAAATATATATLLPPEAPENTSGAGLLEQAAVVEEASD 120

121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVSNNHMSMTN 180
 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVSNNHMSMTN 180
 181 SGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
 181 SGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
 241 SVRYGHRDGGKYAXSGRNGGPA 263
 241 SVRYGHRDGGKYAXSGRNGGPA 263
 RESULT 10
 AAU74591
 ID AAU74591 standard; protein; 358 AA.
 XX
 AC AAU74591;
 XX
 XX 29-AUG-2003 (revised)
 DT 08-MAY-2002 (first entry)
 DT
 XX
 DE Antigenic fusion protein TbH9-Tb38-1.
 XX
 KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; TbH9-Tb38-1; TbH9; Tb38-1.
 XX
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 254 /label= unknown
 FT
 XX US2002009459-A1.
 XX 24-JAN-2002.
 XX
 XX 07-APR-1999; 99US-00287849.
 XX
 XX 13-MAR-1997; 97US-00818112.
 PR 01-OCT-1997; 97US-00342578.
 PR 18-FEB-1998; 98US-00025197.
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 XX (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMP/) CAMPOS-NETO A.
 XX
 XX Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 WPI; 2002-171134/22.
 XX
 XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 XX diagnosing, treating or preventing M. tuberculosis infection,
 XX particularly as vaccine for treating or preventing tuberculosis.
 XX
 PS Claim 1; Fig 4C-D; 62pp; English.
 XX
 CC The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention. Note: The specification states that this polypeptide is
 CC encoded by the polynucleotide shown in ABK14131. (Updated on 29-AUG-2003

XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64503.
 XX
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 XX Example 3B; Page 128-129; 230pp; English.
 PS
 PS This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunisation against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis
 XX
 XX Sequence 391 AA;
 SQ

Query Match 90.9%; Score 1187; DB 2; Length 391;
 Best Local Similarity 99.6%; Pred. No. 8.2e-95;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA	60
Db	74	VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA	133
QY	61	IADVNEAYGEMWAQDAAMFGYAAATATATATATLLPPEAPEMTSAGGLLEQAAAVEEASD	120
Db	134	IADVNEAYGEMWAQDAAMFGYAAATATATATATLLPPEAPEMTSAGGLLEQAAAVEEASD	193
QY	121	TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVMANNHMSMTN	180
Db	194	TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVMANNHMSMTN	253
QY	181	SGVSMNTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSGLGGGVAANLGRAA	240
Db	254	SGVSMNTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSGLGGGVAANLGRAA	313
QY	241	SV 242	
Db	314	SV 315	

RESULT 15

AA04778
 ID AAY04778 standard; protein; 391 AA.
 XX
 AC AAY04778;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species protein sequence 5R.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 OS Mycobacterium sp.
 XX
 PN WO909186-A2.
 XX
 PD 25-FEB-1999.
 XX
 XX 14-AUG-1998; 98WO-FR001813.
 PF
 XX 14-AUG-1997; 97FR-00010404.
 PR
 XX 11-SEP-1997; 97FR-00011325.
 XX
 PA (INSP) INST PASTEUR.
 XX
 XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 PI Cognet De La Salmoniere Y;

DR WPI; 1999-181045/15.
 DR N-PSDB; AAX34030.
 XX
 XX Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX
 XX Claim 32; Fig 5R; 309pp; French.
 PS
 PS Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 XX Sequence 391 AA;
 SQ

Query Match 90.9%; Score 1187; DB 2; Length 391;
 Best Local Similarity 99.6%; Pred. No. 8.2e-95;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA	60
Db	74	VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA	133
QY	61	IADVNEAYGEMWAQDAAMFGYAAATATATATATLLPPEAPEMTSAGGLLEQAAAVEEASD	120
Db	134	IADVNEAYGEMWAQDAAMFGYAAATATATATATLLPPEAPEMTSAGGLLEQAAAVEEASD	193
QY	121	TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVMANNHMSMTN	180
Db	194	TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVMANNHMSMTN	253
QY	181	SGVSMNTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSGLGGGVAANLGRAA	240
Db	254	SGVSMNTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSILGSSGLGGGVAANLGRAA	313
QY	241	SV 242	
Db	314	SV 315	

Search completed: June 30, 2004, 16:48:52
 Job time : 28.0437 secs

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very important document, as it contains the President's annual message to Congress. The letter is written in a formal, dignified style, and it is one of the most important documents in the history of the United States.

2. The second part of the document is a report from the Secretary of the Treasury, dated January 3, 1862. It is a very important document, as it contains the Secretary's annual report to Congress. The report is written in a formal, dignified style, and it is one of the most important documents in the history of the United States.

3. The third part of the document is a report from the Secretary of the Interior, dated January 3, 1862. It is a very important document, as it contains the Secretary's annual report to Congress. The report is written in a formal, dignified style, and it is one of the most important documents in the history of the United States.

4. The fourth part of the document is a report from the Secretary of the Navy, dated January 3, 1862. It is a very important document, as it contains the Secretary's annual report to Congress. The report is written in a formal, dignified style, and it is one of the most important documents in the history of the United States.

5. The fifth part of the document is a report from the Secretary of the War, dated January 3, 1862. It is a very important document, as it contains the Secretary's annual report to Congress. The report is written in a formal, dignified style, and it is one of the most important documents in the history of the United States.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 8.0043 Seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796C-26

Perfect score: 1306

Sequence: 1 VAMSVTAQAEALTAQVRV.....YGRDGGKYXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:**
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/6C COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/6D COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	1304	99.8	263	3	US-08-818-112-91
2	1304	99.8	263	4	US-08-818-111-92
3	1304	99.8	263	4	US-09-056-556-91
4	1304	99.8	263	4	US-09-072-596-92
5	1304	99.8	263	4	US-09-072-967-91
6	1304	99.8	358	4	US-09-287-849-8
7	1187	90.9	391	3	US-08-818-112-107
8	1187	90.9	391	4	US-08-818-111-102
9	1187	90.9	391	4	US-09-056-556-107
10	1187	90.9	391	4	US-09-072-596-102
11	1187	90.9	391	4	US-09-072-967-107
12	1187	90.9	596	4	US-09-287-849-26
13	1187	90.9	600	4	US-09-287-849-22
14	1182	90.5	729	4	US-09-223-040-2
15	1182	90.5	729	4	US-09-287-849-2
16	1001.5	76.7	396	3	US-08-818-112-111
17	1001.5	76.7	396	4	US-08-818-111-106
18	1001.5	76.7	396	4	US-09-056-556-111
19	1001.5	76.7	396	4	US-09-072-596-106
20	1001.5	76.7	396	4	US-09-072-967-111
21	949.5	72.7	359	3	US-08-818-112-109
22	949.5	72.7	359	4	US-08-818-111-104
23	949.5	72.7	359	4	US-09-056-556-109
24	949.5	72.7	359	4	US-09-072-596-104
25	949.5	72.7	359	4	US-09-072-967-109
26	429.5	32.9	400	4	US-09-073-009-126
27	429.5	32.9	400	4	US-09-073-010-126

28 325.5 24.9 423 4 US-09-073-009-142 Sequence 142, App
29 325.5 24.9 423 4 US-09-073-010-142 Sequence 142, App
30 325.5 24.9 710 4 US-09-287-849-16 Sequence 16, Appl
31 325.5 24.9 856 4 US-09-287-849-12 Sequence 12, Appl
32 320.5 24.5 943 4 US-09-477-135A-131 Sequence 131, App
33 239 18.3 204 4 US-08-311-731A-57 Sequence 57, Appl
34 235 18.0 208 4 US-08-311-731A-208 Sequence 208, App
35 223 17.1 141 4 US-09-073-009-15 Sequence 15, Appl
36 223 17.1 141 4 US-09-073-010-15 Sequence 15, Appl
37 197.5 15.1 943 4 US-09-056-556-204 Sequence 204, App
38 197.5 15.1 943 4 US-09-072-596-199 Sequence 199, App
39 197.5 15.1 943 4 US-09-072-967-204 Sequence 204, App
40 188.5 14.4 368 3 US-08-818-112-114 Sequence 114, App
41 188.5 14.4 368 4 US-08-818-111-109 Sequence 109, App
42 188.5 14.4 368 4 US-09-056-556-114 Sequence 114, App
43 188.5 14.4 368 4 US-09-072-596-109 Sequence 109, App
44 188.5 14.4 368 4 US-09-072-967-114 Sequence 114, App
45 188.5 14.4 371 4 US-09-050-739-92 Sequence 92, Appl

ALIGNMENTS

RESULT 1
US-08-818-112-91
; Sequence 91, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J. 31,392
; REGISTRATION NUMBER: 210121.411C6
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-91

Query Match 99.8%; Score 1304; DB 3; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116; Indels 0; Gaps 0;
Matches 263; Conservative 0; Mismatches 0;

Db 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Qy 181 SGVSMNTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLSGSSLGSGGVAANLGRAA 240
Db 181 SGVSMNTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLSGSSLGSGGVAANLGRAA 240
Qy 241 SVRYGHRDGGKYAXSGRRNGGPA 263
Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 4

US-09-072-596-92
; Sequence 92, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-92

Query Match 99.8%; Score 1304; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILLIATNLLGQNTPA 60
Db 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILLIATNLLGQNTPA 60
Qy 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
Db 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
Qy 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Db 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180

Qy 181 SGVSMNTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLSGSSLGSGGVAANLGRAA 240
Db 181 SGVSMNTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLSGSSLGSGGVAANLGRAA 240
Qy 241 SVRYGHRDGGKYAXSGRRNGGPA 263
Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 5

US-09-072-967-91
; Sequence 91, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-91

Query Match 99.8%; Score 1304; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILLIATNLLGQNTPA 60
Db 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILLIATNLLGQNTPA 60
Qy 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
Db 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
Qy 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Db 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAAAQVQTAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
Db |||||
QY 181 SGVSMNTLSSMLKGFAPAAAAAQVQTAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
Db |||||
QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
Db |||||
QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
Db |||||

RESULT 6

US-09-287-849-8
; Sequence 8, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein Tb9-Tb38-1
; NAME/KEY: MOD_RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-09-287-849-8

Query Match 99.8%; Score 1304; DB 4; Length 358;
Best Local Similarity 100.0%; Pred. No. 7.4e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAIATNLGQNTPA 60
Db |||||
QY 1 VAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAIATNLGQNTPA 60
Db |||||
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 120
Db |||||
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 120
Db |||||
QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGLGKWTVPSPHSPISNVSMANNHMSMTN 180
Db |||||
QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGLGKWTVPSPHSPISNVSMANNHMSMTN 180
Db |||||
QY 181 SGVSMNTLSSMLKGFAPAAAAAQVQTAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
Db |||||
QY 181 SGVSMNTLSSMLKGFAPAAAAAQVQTAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
Db |||||
QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
Db |||||
QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
Db |||||

RESULT 7

US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-107

Query Match 90.9%; Score 1187; DB 3; Length 391;
Best Local Similarity 99.6%; Pred. No. 1.1e-104;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAIATNLGQNTPA 60
Db |||||
QY 74 VAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAIATNLGQNTPA 133
Db |||||
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 120
Db |||||
QY 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 193
Db |||||
QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGLGKWTVPSPHSPISNVSMANNHMSMTN 180
Db |||||
QY 194 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGLGKWTVPSPHSPISNVSMANNHMSMTN 253
Db |||||
QY 181 SGVSMNTLSSMLKGFAPAAAAAQVQTAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
Db |||||
QY 254 SGVSMNTLSSMLKGFAPAAAAAQVQTAQNGVRAMSSLSGSSGLGGVAAANLGRAA 313
Db |||||
QY 241 SV 242
Db |||||
QY 314 SV 315
Db |||||

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
US-09-287-849-26

Query Match 90.9%; Score 1187; DB 4; Length 596;
Best Local Similarity 99.6%; Pred. No. 2e-104;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAITNLLGQNTPA 60
DB 82 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAITNLLGQNTPA 141
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 142 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 201
QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVSMANNHSMTN 180
DB 202 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVSMANNHSMTN 261
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAONGVRAMSSILGSSGLGGGVAANLGRAA 240
DB 262 SGVSMNTLSSMLKGFAPAAAQAVQTAONGVRAMSSILGSSGLGGGVAANLGRAA 321
QY 241 SV 242
DB 322 SV 323

RESULT 13
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556

; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-22

Query Match 90.9%; Score 1187; DB 4; Length 600;
Best Local Similarity 99.6%; Pred. No. 2e-104;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAITNLLGQNTPA 60
DB 82 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAITNLLGQNTPA 141
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 142 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 201
QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVSMANNHSMTN 180
DB 202 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVSMANNHSMTN 261
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAONGVRAMSSILGSSGLGGGVAANLGRAA 240
DB 262 SGVSMNTLSSMLKGFAPAAAQAVQTAONGVRAMSSILGSSGLGGGVAANLGRAA 321
QY 241 SV 242
DB 322 SV 323

RESULT 14
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-223-040-2

Query Match 90.5%; Score 1182; DB 4; Length 729;
Best Local Similarity 99.2%; Pred. No. 8.1e-104;
Matches 240; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAITNLLGQNTPA 60
DB 215 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAITNLLGQNTPA 274
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 275 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 334

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:52:58 ; Search time 20.5081 Seconds
(without alignments)
3625.462 Million cell updates

Title: US-09-597-796C-26
Perfect score: 1306
Sequence: 1 VAMMSVTAQGAELTAQVRV.....YGHRDGGKYAXSGRRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

```
Minimum DB seq length: 0
Maximum DB seq length: 200000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp:
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp:
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp:
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp:
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp:
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp:
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query #	Match			Length	DB	ID	Description
		Score	Match	Length				
1	1304	99.8	263	12	US-09-886-349A-12	Sequence 12, Appl		
2	1304	99.8	263	14	US-10-193-002-92	Sequence 92, Appl		
3	1304	99.8	263	14	US-10-084-843-91	Sequence 91, Appl		
4	1304	99.8	263	14	US-10-098-732A-12	Sequence 12, Appl		
5	1304	99.8	358	9	US-09-287-849-8	Sequence 8, Appl		
6	1304	99.8	358	14	US-10-359-460-8	Sequence 8, Appl		
7	1187	90.9	391	12	US-09-886-349A-14	Sequence 14, Appl		
8	1187	90.9	391	14	US-10-193-002-102	Sequence 102, Appl		
9	1187	90.9	391	14	US-10-084-843-107	Sequence 107, Appl		
10	1187	90.9	391	14	US-10-098-732A-14	Sequence 14, Appl		
11	1187	90.9	596	9	US-09-287-849-26	Sequence 26, Appl		
12	1187	90.9	596	12	US-09-886-349A-20	Sequence 20, Appl		
13	1187	90.9	596	14	US-10-359-460-26	Sequence 26, Appl		
14	1187	90.9	596	14	US-10-098-732A-20	Sequence 20, Appl		
15	1187	90.9	600	9	US-09-287-849-22	Sequence 22, Appl		

ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Sequence 22, Appl
Sequence 2, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 21, Appl
Sequence 21, Appl
Sequence 22, Appl
Sequence 15, Appl
Sequence 14, Appl
Sequence 13, Appl
Sequence 65, Appl
Sequence 12, Appl
Sequence 4, Appl
Sequence 18, Appl
Sequence 17, Appl
Sequence 16, Appl
Sequence 2, Appl
Sequence 16, Appl
Sequence 2, Appl
Sequence 16, Appl
Sequence 2, Appl
Sequence 8, Appl
Sequence 106, App
Sequence 111, App
Sequence 104, App
Sequence 109, App
Sequence 62455, A
Sequence 62455, A
Sequence 64892, A
Sequence 126, App
Sequence 126, App
Sequence 62027, A

QY 61 IAVNEAEYGEWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEREASD 120
 Db 61 IAVNEAEYGEWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEREASD 120
 QY 121 TAAANQLMNNVPOALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
 Db 121 TAAANQLMNNVPOALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
 QY 181 SGVSMNTLSSMLKGFAPAAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 Db 181 SGVSMNTLSSMLKGFAPAAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 2

US-10-193-002-92
 ; Sequence 92, Application US/10193002
 ; Publication No. US20030135026A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 Skeiky, Yasir A.W.
 Dillon, Davin C.
 Campos-Neto, Antonia
 Houghton, Raymond
 Vedvick, Thomas S.
 Twardzik, Daniel R.
 Lodes, Michael J.
 Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 TUBERCULOSIS

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/193,002

FILING DATE: 10-Jul-2002

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 92:

SEQUENCE CHARACTERISTICS:

LENGTH: 263 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 92:

US-10-193-002-92

Query Match 99.8%; Score 1304; DB 14; Length 263;
 Best Local Similarity 100.0%; Pred. No. 1.1e-106;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEILTAQVRAAAAYETAYGLTVPPVIAENRAELMILITATLLQONTPA 60
 Db 1 VAMSVTAGQAEILTAQVRAAAAYETAYGLTVPPVIAENRAELMILITATLLQONTPA 60
 QY 61 IAVNEAEYGEWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEREASD 120
 Db 61 IAVNEAEYGEWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEREASD 120
 QY 121 TAAANQLMNNVPOALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
 Db 121 TAAANQLMNNVPOALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
 QY 181 SGVSMNTLSSMLKGFAPAAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 Db 181 SGVSMNTLSSMLKGFAPAAAAQAQVOTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 3

US-10-084-843-91

; Sequence 91, Application US/10084843

; Publication No. US20030143243A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 Skeiky, Yasir A.W.
 Dillon, Davin C.
 Campos-Neto, Antonio
 Houghton, Raymond
 Vedvick, Thomas S.
 Twardzik, Daniel R.
 Lodes, Michael J.
 Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/084,843

FILING DATE: 25-Feb-2002

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967

FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 91:

SEQUENCE CHARACTERISTICS:

LENGTH: 263 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 91:

US-10-084-843-91

Query Match 99.8%; Score 1304; DB 14; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.1e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILLIATNLLGQNTPA 60
DB 1 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILLIATNLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120

QY 121 TAAANQLMNNVPOALKOLAQTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
DB 121 TAAANQLMNNVPOALKOLAQTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 4
US-10-098-732A-12
; Sequence 12, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB39 (TbH9)
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-10-098-732A-12

Query Match 99.8%; Score 1304; DB 14; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.1e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILLIATNLLGQNTPA 60
DB 1 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILLIATNLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120

QY 121 TAAANQLMNNVPOALKOLAQTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
DB 121 TAAANQLMNNVPOALKOLAQTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240

DB 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 5
US-09-287-849-8
; Sequence 8, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neco, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
; OTHER INFORMATION: protein TbH9-Tb38-1
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-09-287-849-8

Query Match 99.8%; Score 1304; DB 9; Length 358;
Best Local Similarity 100.0%; Pred. No. 1.7e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILLIATNLLGQNTPA 60
DB 1 VAWMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILLIATNLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120

QY 121 TAAANQLMNNVPOALKOLAQTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
DB 121 TAAANQLMNNVPOALKOLAQTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/10/193,002
 ; FILING DATE: 10-Jul-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,596
 ; FILING DATE: 05-MAY-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 102:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 102:
 ; US-10-193-002-102

Query Match 90.9%; Score 1187; DB 14; Length 391;
 Best Local Similarity 99.6%; Pred. No. 3.8e-96;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 74 VAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 133
 QY 61 IAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPPEEAPEMTSAGLLLEQAAAVEEASD 120
 DB 134 IAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPPEEAPEMTSAGLLLEQAAAVEEASD 193
 QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPHSPI-SNMVSMANNHSMTN 180
 DB 194 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPHSPI-SNMVSMANNHSMTN 253
 QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
 DB 254 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSGLGGGVAANLGRAA 313
 QY 241 SV 242
 DB 314 SV 315

RESULT 9
 US-10-084-843-107
 ; Sequence 107, Application US/10084843
 ; Publication No. US20030143243A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonio
 ; Houghton, Raymond
 ; Vedwick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 ; AND DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 355
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington

; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/084,843
 ; FILING DATE: 25-Feb-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,967
 ; FILING DATE: 05-MAY-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 107:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 107:
 ; US-10-084-843-107

Query Match 90.9%; Score 1187; DB 14; Length 391;
 Best Local Similarity 99.6%; Pred. No. 3.8e-96;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 74 VAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 133
 QY 61 IAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPPEEAPEMTSAGLLLEQAAAVEEASD 120
 DB 134 IAVNEAEYGEWMAQDAAMFGVAAAATATATATLLPPEEAPEMTSAGLLLEQAAAVEEASD 193
 QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPHSPI-SNMVSMANNHSMTN 180
 DB 194 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPHSPI-SNMVSMANNHSMTN 253
 QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
 DB 254 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSGLGGGVAANLGRAA 313
 QY 241 SV 242
 DB 314 SV 315

RESULT 10
 US-10-098-732A-14
 ; Sequence 14, Application US/10098732A
 ; Publication No. US20030175294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; Applicant: Brannon, Mark
 ; Applicant: Guderian, Jeffrey
 ; Applicant: Corixa Corporation
 ; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; FILE OF INVENTION: Leishmania Antigen
 ; FILE REFERENCE: 014058-012010US
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: US 60/275,837
 ; PRIOR FILING DATE: 2001-03-13
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB39 full length (TBH9FL)
US-10-038-732A-14

Query Match 90.9%; Score 1187; DB 14; Length 391;
Best Local Similarity 99.6%; Pred. No. 3.8e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAGQAEELTAQVRVAAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA	60
DB	74	VAMSVTAGQAEELTAQVRVAAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA	133
QY	61	IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD	120
DB	134	IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD	193
QY	121	TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN	180
DB	194	TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN	253
QY	181	SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA	240
DB	254	SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA	313
QY	241	SV 242	
DB	314	SV 315	

RESULT 11
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 90.9%; Score 1187; DB 9; Length 596;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
DB 82 VAMSVTAGQAEELTAQVRVAAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 141
QY 61 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 142 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 201
QY 121 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
DB 202 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 261
QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 262 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 321
QY 241 SV 242
DB 322 SV 323

RESULT 12
US-09-886-349A-20
; Sequence 20, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TbH9-Ra35 (designated MTB59F)
US-09-886-349A-20

Query Match 90.9%; Score 1187; DB 12; Length 596;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
DB 82 VAMSVTAGQAEELTAQVRVAAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 141
QY 61 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 142 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 201
QY 121 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
DB 202 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 261
QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 262 SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 321
QY 241 SV 242
DB 322 SV 323


```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-22

Query Match          90.9%; Score 1187; DB 9; Length 600;
Best Local Similarity 99.6%; Pred. No. 6.7e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAATNLLGQNTPA 60
Db 82 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAATNLLGQNTPA 141

QY 61 IAVNEAEGEMWAQDAAMFGYAAATATATATILPPEAPENTSAGGLLEQAAVEEASD 120
Db 142 IAVNEAEGEMWAQDAAMFGYAAATATATATILPPEAPENTSAGGLLEQAAVEEASD 201

QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLMKTVPSPHRSPISNMVMANNHMSMTN 180
Db 202 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLMKTVPSPHRSPISNMVMANNHMSMTN 261

QY 181 SGVSMNTLSSMLKGFAPAAAQAVCTAAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
Db 262 SGVSMNTLSSMLKGFAPAAAQAVCTAAQNGVRAMSSLGSSGLGGGVAANLGRAA 321

QY 241 SV 242
Db 322 SV 323
```

Search completed: June 30, 2004, 17:14:49
Job time : 21.5081 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 6.87361 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796C-26
Perfect score: 1306
Sequence: 1 VAMSVTAGQBELTAAQVRV.....YGHRDGKXYSGRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1187	90.9	391	2 B70608	probable PPE prote
2	1001.5	76.7	396	2 H70741	probable PPE prote
3	949.5	72.7	393	2 C70568	probable PPE prote
4	460	35.2	403	2 H70931	probable PPE prote
5	441	33.8	423	2 B70931	probable PPE prote
6	432.5	33.1	393	2 G70929	probable PPE prote
7	424.5	32.5	421	2 H87056	PPE-family protein
8	422	32.3	409	2 A70932	probable PPE prote
9	403	30.9	413	2 F70560	probable PPE prote
10	402	30.8	391	2 B70625	probable PPE prote
11	402	30.8	463	2 C70931	probable PPE prote
12	395	30.2	408	2 G70925	probable PPE prote
13	389	29.8	394	2 G70881	probable PPE prote
14	386.5	29.6	468	2 A70932	probable PPE prote
15	379	29.0	380	2 B70646	probable PPE prote
16	375	28.7	385	2 H70503	probable PPE prote
17	352	27.0	402	2 A70882	probable PPE prote
18	348.5	26.7	365	2 G70929	probable PPE prote
19	346.5	26.5	394	2 A70504	probable PPE prote
20	346	26.5	350	2 H70929	probable PPE prote
21	339	25.2	391	2 D70922	probable PPE prote
22	328	25.1	406	2 E70675	probable PPE prote
23	325.5	24.9	423	2 C70582	probable PPE prote
24	320.5	24.5	3300	2 D70575	probable PPE prote
25	317	24.3	391	2 A70663	probable PPE prote
26	282	21.6	963	2 B70524	probable PPE prote
27	272	20.8	580	2 G70570	probable PPE prote
28	271.5	20.8	3716	2 E70969	probable PPE prote
29	266	20.4	180	2 G70834	probable PPE prote

30	257	19.7	618	2 H70552	probable PPE prote
31	255.5	19.6	678	2 A70762	probable PPE prote
32	250.5	19.2	1053	2 E70987	probable PPE prote
33	245	18.8	615	2 E70663	probable PPE prote
34	243.5	18.6	443	2 C70780	probable PPE prote
35	243	18.6	487	2 C70830	probable PPE prote
36	242.5	18.6	346	2 H70874	probable PPE prote
37	242	18.5	2523	2 F70846	probable PPE prote
38	239	18.3	655	2 A70931	probable PPE prote
39	238	18.2	3157	2 E70969	probable PPE prote
40	236	18.1	479	2 D70676	probable PPE prote
41	235	18.0	204	2 T45168	probable PPE prote
42	231.5	17.7	987	2 E70808	probable PPE prote
43	229	17.5	1436	2 B70520	probable PPE prote
44	227	17.4	582	2 F70675	probable PPE prote
45	223.5	17.1	645	2 F70825	probable PPE prote

ALIGNMENTS

RESULT 1

B70608
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70608
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:9829598; PMID:9634230
A:Accession: B70608
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-391 <COL>
A:Cross-references: GB:Z93777; GB:AL123456; NID:G3261726; PIDN:CAB07839.1; PID:e311073; I
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 90.9%; Score 1187; DB 2; Length 391;
Best Local Similarity 99.6%; Pred.No. 8.6e-74;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMILIAITNLLGQNTPA	60
DB	74	VAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMILIAITNLLGQNTPA	133
QY	61	I AVNEAYGENWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGLLFOAAVEEASD	120
DB	134	I AVNEAYGENWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGLLFOAAVEEASD	193
QY	121	TAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN	180
DB	194	TAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN	253
QY	181	SGVSMNTTLLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSSLGSSGLGGVAAANLGRAA	240
DB	254	SGVSMNTTLLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSSLGSSGLGGVAAANLGRAA	313
QY	241	SV 242	
DB	314	SV 315	

RESULT 2

H70741
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70741
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: H70741
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-396 <COL>
 A: Cross-references: GB:275555; GB:AL123456; NID:G3261608; PIDN:CAA99966.1; PID:e250360;
 A: Experimental source: strain H37Rv
 C: Genetics:
 A: Gene: PPE

Query Match 76.7%; Score 1001.5; DB 2; Length 396;
 Best Local Similarity 84.1%; Pred. No. 3.9e-61;
 Matches 207; Conservative 13; Mismatches 21; Indels 5; Gaps 2;
 QY 1 VAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 60
 DB 74 VAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 133
 QY 61 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGGLLEQAAVVEASD 120
 DB 134 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGGLLEQAAVVEASD 193
 QY 121 TAAANQLMNNVPQALQKQAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 180
 DB 194 TAAANQLMNNVPQALQKQAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 253
 QY 181 SGVSWNTLTSLMLKGFAPAAAQAQVOTAAONGVAMSS---LGSSLGSSGLGGVAAANL 236
 DB 254 SGVSWNTLTSLMLKGFAPAAAQAQVOTAAONGVAMSSLGSSLGSSGLGGVAAANL 312
 QY 237 GRAASV 242
 DB 313 GRAASV 318

RESULT 3
 C70568
 probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C: Accession: C70568
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: C70568
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-393 <COL>
 A: Cross-references: GB:295390; GB:AL123456; NID:G3261766; PIDN:CAB08702.1; PID:e316074;
 A: Experimental source: strain H37Rv
 C: Genetics:
 A: Gene: PPE

Query Match 72.7%; Score 949.5; DB 2; Length 393;
 Best Local Similarity 79.7%; Pred. No. 1.4e-57;
 Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;
 QY 1 VAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 60
 DB 74 VAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 133
 QY 61 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGGLLEQAAVVEASD 120

Db 134 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGGLLEQAAVVEASD 193
 QY 121 TAAANQLMNNVPQALQKQAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 180
 Db 194 TAAANQLMNNVPQALQKQAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 253
 QY 181 SGVSWNTLTSLMLKGFAPAAAQAQVOTAAONGVAMSS---LGSSLGSSGLGGVAAANL 236
 Db 254 SGVSWNTLTSLMLKGFAPAAAQAQVOTAAONGVAMSSLGSSLGSSGLGGVAAANL 312
 QY 237 GRAASV 242
 Db 313 GRAASV 318

RESULT 4
 H70931
 probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C: Accession: H70931
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: H70931
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-403 <COL>
 A: Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN:CAA17728.1; PID:e125461;
 A: Experimental source: strain H37Rv
 C: Genetics:
 A: Gene: PPE

Query Match 35.2%; Score 460; DB 2; Length 403;
 Best Local Similarity 43.3%; Pred. No. 3.4e-24;
 Matches 117; Conservative 36; Mismatches 73; Indels 44; Gaps 8;
 QY 1 VAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 60
 Db 77 VAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 136
 QY 61 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGGLLEQAAVVEASD 120
 Db 137 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGGLLEQAAVVEASD 195
 QY 121 TAAANQLMNNVPQALQKQAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 161
 Db 196 ASAGAQQTTLSQIAAIPSVLQGLSSSTAAATFASGGLGIVGSGSWLDKMLWALDPN 255
 QY 162 RSPISNMVSMANNHSMNTNNGVSM-TNTLS---SNLKGFAFAAAAQAVQTAQNGVAMSS 217
 Db 256 -----SNFWNTIASSGLFLPSNTIAPFLGLGGVAAADAAGDVLGEATSG----- 300
 QY 218 SIGSSL-----GSSGLGGVAAANTGRAASV 242
 Db 301 GLGGALVAPLGSAGGLGCTVAAGLGNATV 330

RESULT 5
 B70931
 probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C: Accession: B70931
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998

A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70931
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-423 <COL>
A;Cross-references: GB:AL020201; GB:AL123456; NID:g3250699; PIDN:CAA1772.1; PID:e125461
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: PPE

Query Match 33.1%; Score 432.5; DB 2; Length 393;
Best Local Similarity 41.0%; Pred. No. 2.4e-22; Indels 31; Gaps 7;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
C;Species: Mycobacterium tuberculosis (strain H37RV)
C;Date: 17-Jul-1998 #sequence_revision 38; Mismatches 85; Indels 31; Gaps 7;
C;Accession: G70929
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70929
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-393 <COL>
A;Cross-references: GB:AL020201; GB:AL123456; NID:g3250699; PIDN:CAA1771.1; PID:e125460
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: PPE

Query Match 32.5%; Score 424.5; DB 2; Length 421;
Best Local Similarity 39.8%; Pred. No. 9.3e-22; Indels 29; Gaps 5;
Matches 104; Conservative 36; Mismatches 99; Indels 29; Gaps 5;
C;Species: Mycobacterium tuberculosis (strain H37RV)
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: G70929
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70929
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-421 <STO>
A;Cross-references: GB:AL450380; NID:g13093150; PIDN:CAC31563.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML1182

Query Match 33.1%; Score 432.5; DB 2; Length 393;
Best Local Similarity 41.0%; Pred. No. 2.4e-22; Indels 31; Gaps 7;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
C;Species: Mycobacterium tuberculosis (strain H37RV)
C;Date: 17-Jul-1998 #sequence_revision 38; Mismatches 85; Indels 31; Gaps 7;
C;Accession: G70929
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70929
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-393 <COL>
A;Cross-references: GB:AL020201; GB:AL123456; NID:g3250699; PIDN:CAA1771.1; PID:e125460
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: PPE

Db 192 TAQSTLTEMITGPNALQSLTSPILQSS-NGPLSLWMLQILFGTFNFTSIALLTDLQPY 250
QY 162 RSPISNMVSMANNHSMNTSGVMTNTLSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSIGS 221
Db 251 ASFYNTTEGLPYFSGNNGNFIQSAKTL-GLIGSAPAAVA-----AAGDAKGLPGLGG 304
QY 222 SLGSSGLGGGVAANLGRAASV 242
Db 305 MLG----GGPVAAGLGRAASV 321

RESULT 7
H87056
PPE-family protein [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: H87056
R;Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: H87056
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-421 <STO>
A;Cross-references: GB:AL450380; NID:g13093150; PIDN:CAC31563.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML1182

Query Match 32.5%; Score 424.5; DB 2; Length 421;
Best Local Similarity 39.8%; Pred. No. 9.3e-22; Indels 29; Gaps 5;
Matches 104; Conservative 36; Mismatches 99; Indels 29; Gaps 5;
C;Species: Mycobacterium tuberculosis (strain H37RV)
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: G70932
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70932
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

Query Match 33.1%; Score 432.5; DB 2; Length 393;
Best Local Similarity 41.0%; Pred. No. 2.4e-22; Indels 31; Gaps 7;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
C;Species: Mycobacterium tuberculosis (strain H37RV)
C;Date: 17-Jul-1998 #sequence_revision 38; Mismatches 85; Indels 31; Gaps 7;
C;Accession: G70929
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70932
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

Query Match	30.8%	Score	402;	DB	2;	Length	463;	
Best Local Similarity	35.7%	Pred.	No. 3.6e-20;					
Matches	114;	Conservative	37;	Mismatches	106;	Indels	62; Gaps	8;

Qy	1	VAMSVTAGQAELTAAGVRVAAAAYETAYGLTVPPTVAENRAEMLMILLIATNLGQNTPA	60
Dd	72	VAMSATAALAREAAQASAAAAAYEAFAATVPFPVVANRAELAVLAATNIPGQTGA	131
Qy	61	IATNEAEEYGEWMAQDAAMFGYAATAATATATLLPFEEAPEMTSAGLLEQAABEASD	120
Dd	132	IAAEARYAEMWAQDAAMAAMYGYAGSSSVAT-QVTFFAAPPPTTNAAGLTQGVAVAQVG	190
Qy	121	TAAANQLMNVPQALKQLAQDTQGTTFSSKLGGLWKTVS--PHRSP I-----SN	167
Dd	191	ASAGN-ARSLYSEVLEFLA--TAGTNYNKTVASLMNAVTPGVFYASSVYNSMLGLGF AESK	247
Qy	168	MVSMMNNHM-----SMTNSGVSMINTLSMKLGKGPAA	200
Dd	248	NVLPA NDTVISITFGMVQFQKFNPVPFNPDLPKSGALGAGLGRSAISSGLGSTAPAI	307
Qy	201	AAQAVQTAQNQNV R-----AMSLGSLSGSSGLGGVAAANLGRA-----ASV	242
Dd	308	SAGASQAGSVGMSVPPSWAAATPAIRTVAAVFSTGLQAVPAAAISEGSI LLSQMALASV	367
Qy	243	RYGHRDGGKYAXSGRRNGG	261
Dd	368	AGGALGGAARAATGCGFLGG	386

RESULT 12
G70925

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: G70925
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: G70925
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-408 <COL>
A/Cross-references: GB:Z74024; GB:AL123456; NID:93250700; PIDN:CAA96377.1; PID:ei1301025;
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: ppe

Query Match	30.2%;	Score 395;	DB 2;	Length 408;
Best Local Similarity	38.9%;	Pred. No. 9.2-20;		
Matches 110; Conservative 31;	Mismatches 78;	Indels 64;	Gaps 11;	
Qy	1	VAMSVTAGQAELTAAGVRVAAAAYETAYGLTVPVPPVIAENRAELMILATLNLGONTPA	60	
Dd.	73	VAMLSATAGAQRGMQARAAAAAYELAFANTVPPVVVANRALLVALVATNFGONTPA	132	
Qy	61	IANNAEYGENWAQDAAMFCYAAATATATATLLPFEAPEMTSAGLLSQAAA-----V	115	
Dd	133	IAATEAQYAEMWADAAAMYAYAGSAAIAT--ELTFPTAAPVTTPSALAGQAATAVSSTV	191	
Qy	116	EASDTHAANLOMN-----VP--QALKO-LAOPTOGTTTSSKL-----GGL	154	
Dd	192	PPLATTAAVPQLQLSSTSLLIPWSALQWLAEHLGLTPDNEMTVIRLLGISYPDEGL	251	
Qy	155	WKTVSPHRSPISNVMVMANNMTNGVSMNTLTSSLMLKGFAPAAAAAQVTAQQNGVR	214	
Dd	252	-----LQFRASLAQQAIPCTPGAG--DSGSSVLDSWGPTIFA-----GPR	290	
Qy	215	AMSSL--GSSSLGS-----SGLGCGGVAANLGRAAS	241	

Db 291 ASPSVAGGAVGGVGTPOPYWYWALDRESIGGSVAALKGSS 333

RESULT 13

7G0881

probable PFB protein - Mycobacterium tuberculosis (strain H37Rv)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_channel
C/Accession: G70881

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C/Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton
Nature 393, 537-544, 1998

A/Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead,
A/Title: Deciphering the biology of Mycobacterium tuberculosis
A/Reference number: A70500; MUID:9825987; PMID:9634230

A/Accession: G70881

A/Status: preliminary; nucleic acid sequence not shown; trans
A/Molecule type: DNA

A/Residues: 1-394 <COL>

A/Cross-references: GB:AL008967; GB:AL123456; NID:g3261491; P
A/Experimental source: strain H37Rv

C/Genetics:
A/Genes: bpf

Query Match.	29.8%;	Score 389;	DB 2;	Length 394;
Best Local Similarity	37.6%;	Pred.No. 2.3e-19;		
Matches 103; Conservative	43;	Mismatches 106;	Indels	22; Gaps 7;
Qy	1	VAMSVTAGQAETAAQURVAAAAYETAYGLTVPPPIAENRAELMILIIATNLLGQNTTPA	60	
		:::::	:::::	
Db	73	LAWLITYTAEAGAHAAQAQAVASAAFAEFAMTVPEAEVAANRALLAALVAINVLGQNTTPA	132	
		:::::	:::::	
Qy	61	IANYEABYGEWMAQDAAMAEGYAAATATATATLLPPEEAPEMTSACGLLEQAAAAVEASD	120	
		:::::	:::::	
Db	133	IMATEAHYGEWMAQDALAMTYGYAASSA--AAGRNLPLITPQTANMAGLAGQAAAAVSRAAA	191	
		:::::	:::::	
Qy	121	TAAANAQ-----LMNVNPQALKQLAQPTQGTTPTPSKLGGLMKTWSP--HRSPISNMVSMAN	173	
		:::::	:::::	
Db	192	ASTVQQVGLGSLISLNLPNAVGMFPASPLTSADAAGLGGIIOIDIELLGITFVQNAINGAV	251	
		:::::	:::::	
Qy	174	NHMSMTNSGVSYMTNLSMLKGPAAPAAAQAVTAQONGV---RAMSSLGSSLGSSGLGG	230	
		:::::	:::::	
Db	252	N----TTAFVFVNATIPNAVFLGHAFALNFATVTATAADAVFAAAAAAGLAHTVTFVGVGG	307	
		:::::	:::::	
Qy	231	-GVAANLGRAASVRYGRHDGCKYXSGRRNGGPA	263	
		:::::	:::::	
Db	308	ASLTASLGEASSV-----GGLSVPAQGWSTAAPA	335	
		:::::	:::::	

RESULT 14

B70932 probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
B70932 C#Species: Mycobacterium tuberculosis
C#Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C#Accession: B70932
R#Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris,
; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentile, S.; Hamlin, N.;
Rajadaram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares,
Nature 393, 537-544, 1998
A#Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.
A#Title: Deciphering the biology of Mycobacterium tuberculosis from the com
A#Reference number: A70500; MUID:98295987; PMID:9634230
A#Accession: B70932
A#Status: preliminary; nucleic acid sequence not shown; translation not sho
A#Molecule type: DNA
A#Residues: 1-468 <COL>
A#Cross-references: GB:AI020201; GB:AI123456; NTD:93250699; PIDN:CAAL7730.1
A#Experimental source: strain H37RV
C#Genetics:
A#Gene: PPE

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 4.28192 Seconds
(without alignments)
3198.204 Million cell updates/sec

Title: US-09-597-796C-26

Perfect score: 1306

Sequence: 1 VAWMSVTAGQALFRAQVRV.....YCHRDGGKYXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1001.5	76.7	396	1 YD61_MYCTU	Q11031 mycobacteri
2	402	30.8	463	1 Y102_MYCTU	O53951 mycobacteri
3	395	30.2	408	1 YS92_MYCTU	Q10813 mycobacteri
4	255.5	19.6	678	1 YF48_MYCTU	Q10778 mycobacteri
5	243.5	18.6	443	1 Y878_MYCTU	Q10540 mycobacteri
6	243	18.6	487	1 Y442_MYCTU	P42611 mycobacteri
7	194.5	14.9	463	1 Y096_MYCTU	Q10892 mycobacteri
8	191.5	14.7	408	1 SRA_MYCLE	Q07297 mycobacteri
9	158.5	12.1	434	1 YU18_MYCTU	P31500 mycobacteri
10	158	12.1	435	1 YU21_MYCTU	O53268 mycobacteri
11	156	11.9	178	1 YY29_MYCTU	O06246 mycobacteri
12	152	11.6	176	1 YY25_MYCTU	Q50703 mycobacteri
13	114.5	8.8	1211	1 BUN2_DROME	Q24523 drosophila
14	112	8.6	2090	1 N214_HUMAN	P35658 homo sapien
15	110.5	8.5	938	1 TRG1_ECOLI	P33790 escherichia
16	107.5	8.2	354	1 YAGU_SCHPO	Q10169 schizosacch
17	106	8.1	1783	1 RAA3_CHLRE	P97ec4 chlamydomyc
18	103.5	7.9	881	1 PRY3_YEAST	P76072 escherichia
19	102	7.8	1120	1 STFR_ECOLI	P32768 saccharomyc
20	100.5	7.7	1537	1 FLO1_YEAST	P36110 saccharomyc
21	99	7.6	329	1 PRV2_YEAST	Q8n2m8 homo sapien
22	99	7.6	659	1 SFRG_HUMAN	Q04893 saccharomyc
23	99	7.6	1140	1 YN96_YEAST	P40472 saccharomyc
24	98.5	7.5	475	1 SIM1_YEAST	Q61572 mus musculus
25	98.5	7.5	553	1 FXC1_MOUSE	Q10873 mycobacteri
26	98.5	7.5	558	1 YJ83_MYCTU	P33240 homo sapien
27	98.5	7.5	577	1 CST2_HUMAN	P14559 homo sapien
28	98.5	7.5	743	1 PO21_HUMAN	P38594 saccharomyc
29	98.5	7.5	1075	1 FLO5_YEAST	Q9ul36 homo sapien
30	98.5	7.5	1845	1 Z236_HUMAN	P28284 herpes simp
31	98	7.5	825	1 ICP0_HSV2H	Q01981 emericella
32	97.5	7.5	416	1 CREA_EMENI	Q63850 mus musculus
33	97.5	7.5	526	1 NU62_MOUSE	

ALIGNMENTS

RESULT 1

YD61_MYCTU STANDARD; PRT; 396 AA.
AC Q11031;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Hypothetical PFB-family protein Rv1361C/MT1406.
GN Rv1361C OR MT1406 OR MTCY02B10.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala E., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22208494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PFB family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z75555; AAA99966.1; --
DR EMBL; AE007013; AAK45669.1; --
DR FIC; H70741; H70741.
DR TIGR; MT1406; --
DR Tuberculist; Rv1361C; --
DR InterPro; IPR000030; Microbac_PPB.

34 96.5 7.4 461 1 US45_LACLC
35 96.5 7.4 461 1 Y514_HUMAN
36 96.5 7.4 632 1 PO21_RAT
37 96 7.4 939 1 SLAP_CAUF6
38 96 7.4 1025 1 SLAP_CAUF6
39 95.5 7.3 2035 1 HFC1_HUMAN
40 95.5 7.3 2090 1 HFC1_MESAU
41 95 7.3 323 1 JUND_CHICK
42 95 7.3 774 1 STF_LAMBD
43 95 7.3 1556 1 PROS_DROVI
44 94.5 7.2 601 1 PDM1_DROME
45 94.5 7.2 780 1 CTPA_MYCLE

P22865 lactococcus
P60259 homo sapien
P31503 rattus norv
P35827 campylobact
P35828 campylobact
P51610 homo sapien
P51611 mesocricetu
P27921 gallus gall
P27921 gallus gall
P03784 bacteriopho
Q96a1 drosophila
P31368 drosophila
P46839 mycobacteri

```

DR Pfam: PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 158 159 TA -> AT (IN REF. 2).
SQ SEQUENCE 396 AA; 40015 MW; 6AFAB0D7B5F668D0 CRC64;

Query Match 76.7%; Score 1001.5; DB 1; Length 396;
Best Local Similarity 84.1%; Pred. No. 2.4e-62;
Matches 207; Conservative 13; Mismatches 21; Indels 5; Gaps 2;

QY 1 VAMSVTAGQAEALTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 60
DB 74 VAMSVTAGQAEALTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 133
QY 61 IAVNEASYGEMWAQDAAMFGYAATATATATLLPPEAEPMTSAGGLLEQAAVEEASD 120
DB 134 IAVNEASYGEMWAQDAAMFGYAATATATATLLPPEAEPMTSAGGLLEQAAVEEASD 193
QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHSMTN 180
DB 194 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHSMTN 253
QY 181 SGVSMATLHSMKGFAP-AAAQVETAAQNGVRAAMS-----LGSSILGSSGLGGVAAANL 236
DB 254 SGVSMATLHSMKGFAP-AAAQVETAAQNGVRAAMS-----LGSSILGSSGLGGVAAANL 312
QY 237 GRAASV 242
DB 313 GRAASV 318

RESULT 2
YI02 MYCTU STANDARD; PRT; 463 AA.
AC 053951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV1851/MT1851/MB1830.
GN RV1851 OR MT1851 OR MT049.24 OR MB1830.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL022021; CAAL17723.1; -
DR EMBL: AE007044; AAK46123.1; -
DR EMBL: BX248340; CAD94533.1; -
DR PIR: C70931; C70931.
DR TIGR: MT1851; -
DR TubercuList; RV1802; -
DR InterPro; IPR00030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 401 401 S -> L (IN REF. 2).
SQ SEQUENCE 463 AA; 46021 MW; EE64828BF09FA551 CRC64;

Query Match 30.8%; Score 402; DB 1; Length 463;
Best Local Similarity 35.7%; Pred. No. 8.2e-21;
Matches 114; Conservative 37; Mismatches 106; Indels 62; Gaps 8;

QY 1 VAMSVTAGQAEALTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 60
DB 72 VAMSVTAGQAEALTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 131
QY 61 IAVNEASYGEMWAQDAAMFGYAATATATATLLPPEAEPMTSAGGLLEQAAVEEASD 120
DB 132 IAAAEARYAEMWAQDAAMFGYAATATATATLLPPEAEPMTSAGGLLEQAAVEEASD 190
QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPHRSPI-----SN 167
DB 191 ASAGN-ARSLVSEVLEFLA--TAGTNTNKTIVASLNNATGVFVYASSVYNSVLMGLGFASK 247
QY 168 MVSMANNH-----SMTNSGVSMNTLSSMLKGFAPAA 200
DB 248 MYLPANDTVISTIFGMVQFQKFNPTFPNPDLPKLSALGGLRSLAISGLGSTAPAI 307
QY 201 AAQAVQTAQNGVR-----AMSSLSGLSSGLGGVAAANLGRA-----ASV 242
DB 308 SAGASQAGSVGMSVPPSMAAATPAIRTAAVFSTGLQAVPAAAISEGLSQAALASV 367
QY 243 RYGRHDSGKXSGRRNGG 261
DB 368 AGGALGGAARATGGLG 386

RESULT 3
YS92 MYCTU STANDARD; PRT; 408 AA.
AC Q10813;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV2892C/MT2959/MB2916C.
GN RV2892C OR MT2959 OR MT2974.23C OR MB2916C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;

```

```

[1].
RN RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feldwell T., Gencies S., Hamlin M., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
[2]
RN RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., Hickey O.,
RA Peterson J., DeBoy R., Dodson R., Gunn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
[3]
RN RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garner T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR ENBL; Z74024; CAA98377.1; -.
DR ENBL; AB007119; AAK47285.1; -.
DR ENBL; BX248344; CAD96603.1; -.
DR PIR; G70925; G70925.
DR TIGR; MT2959; -.
DR TubercuList; Rv2892c; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Transmembrane; Complete proteome.
FW TRANSMEM 56 POTENTIAL.
SQ SEQUENCE 408 AA; 41469 MW; 3E3D1F2D07B27199 CRC64;
Query Match 30.2%; Score 395; DB 1; Length 408;
Best local similarity 38.9%; Pred No. 2.e-20;
Matches 110; Conservative 78; Indels 64; Gaps 11;
Qy 1 VAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPPVIAENRAELMILIATNLGQTTPA 60
Db 73 VAMLSATGAQAEGACMGQARAATAAYELAFAMTVPPPVVVANRALLVAVATNFFGQNTPA 132
Qy 61 IANVEARYGEWADAAAMFGYAATAATATATALLPREEAPEMTSAGLLAQAAA----V 115
Db 133 IAATEAGYAEWAQDAAMTYAGSAAIAT-ELTPTFAAPVTTSPALAQAAATVSSTV 191
Qy 116 EEASDTAAANQLMNN-----VP--QALKQ-LAQPTGTTPSSKL-----GGI 154
SD :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 192 PPIATTAAPVQLLQQLSSSTSLIPWYSGALQOWLAENLLGLTPDNRMTIVIRLLGISYDFDEGL 251
 QY 155 WKTVSPERGPISNNVSMANNHMSMTNSGVSTNTLSSMLKGFAPAAAOAVQTAQNGVR 214
 Db 252 -----LQFEASIAQAIGTPEGAG--DSGSLVDSWGPITFA-----GPR 290
 QY 215 AMSSL--GSLGSG-----SGLGGGVAANILGRAAS 241
 Db 291 ASPSVAGGAGVGQVTPQFPYVWALDRESIGGSVAALGKSS 333
 RESULT 4
 YF48_MYCTU
 ID_YF48_MYCTU STANDARD; PRF; 678 AA.
 AC Q10778;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rvi548c/Mt1599.
 GN Rvi548C OR MT1599 OR MTCY48.17.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STEAIN=H37Rv;
 RX MEDLINE=98255987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Badham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=42206494; PubMed=12218036;
 RA Petersmann J., DeBo R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490 (2002).
 CC -1. SIMILARITY: Belongs to the mycobacterial PPE family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/ebis-sib.ch/announcement>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; Z74020; CAA98335.1; --
 DR EMBL; AE007026; AAK45866.1; ALT_INIT.
 DR PIR; A70762; A70762.
 DR TIGR; MT1599; --
 DR TubercuList; Rvi548c; --
 DR InterPro; IPR000030; Microbac_PPE.
 DR InterPro; IPR002989; Mycobac_Pentapep.
 DR Pfam; PF01469; Pentapeptide_2; 11.
 DR Pfam; PF00823; PPE; 1.
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 180 200 POTENTIAL.
 FT


```

RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15467; AAA88235.1; AUT_INIT.
DR EMBL; AL021932; CAAL7399.1; -.
DR EMBL; A3006948; AAK44681.1; -.
DR PIR; C70830; C70830.
DR TIGR; MT0458; -.
DR TubercuList; RV00458; -.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR002989; Microbac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 5.
DR Pfam; PF00823; PPE; 1.
DR KX Hypothetical protein; Complete proteome.
FT CONFLICT 40 40 E -> X (IN REF. 2).
FT CONFLICT 96 96 I -> T (IN REF. 1).
FT CONFLICT 211 211 G -> GNNNIG (IN REF. 1).
FT SEQUENCE 487 AA; 47247 MW; 97234D5B316C8C7F CRC64;
SQ
Query Match 18.6%; Score 243; DB 1; Length 487;
Best Local Similarity 28.8%; Pred. No. 8.6e-10;
Matches 83; Conservative 30; Mismatches 129; Indels 46; Gaps 8;
QY 1 VAMSVTAGAQLTAQVRVAAAYATAYGLTVPPVPIAENRAELMILLIATNLLGNTFA 60
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 LAWLSTAAQAQEAQAQAQAIAFAFAAALTAQPAVAANRGLMQLLAATNFWGQNAFA 135
QY 61 IAVNEAYEGMVAQDAADAAAFGVAATATATATLPPFEAPEMTSAG-----GLLEQAAA 114
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 LMDVEAAYEQWALDVAAWAGYHFDASAQAQAPQOV--LRNLGIDGKNGQINLFG 193
QY 115 VEEASPTAANQLMNVPPQALKQAQPTQOTTPSSKLG-----GLWKTVS----- 159
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 NTGSGNIGNNNIGNNIGSGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTG 253
QY 160 ---PHRSPTSNVYMANHNSMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNVRAM 216
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ITGDHQMGFGFGSGGN--TGFNGSGNNGVGLFNS-----GSGNIGNIGNS 297
QY 217 SSLGSLGSGSLGGVGAANLGRAVRYGHRDGG--KYAXSGRNGGPA 263
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 GSLNSGIGTSGT---INAGLGSAGSLNTPFNAGMQLAALGSAAGSEA 342
```

RESULT 7

```

Y096 MYCTU STANDARD; PRT; 463 AA.
ID Y096 MYCTU
AC Q10892;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0096/MT0105.
GN RV0096 OR MT0105 OR MTCY251.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74410; CAA98932.1; -.
DR EMBL; AE006922; NAK44327.1; -.
DR TIGR; H70750; H70750.
DR TubercuList; RV0096; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
SQ SEQUENCE 463 AA; 46894 MW; 42D9D66A033D0DD8 CRC64;
Query Match 14.9%; Score 194.5; DB 1; Length 463;
Best Local Similarity 36.3%; Pred. No. 1.8e-06;
```

Matches 53; Conservative 21; Mismatches 55; Indels 17; Gaps 3;
 QY 1 VAWNSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMLIATNLGQNTPA 60
 DB 70 LAWLEQTAINSAVTAQAQVAAAYACALAMPPTFAELAAHAIHGVLIATNFFGINTVP 129
 QY 61 IAVNEAEYGEWMAODAAAMFYAAATATATATLLPFEAEPMTSAGLLLEQAAAVEASD 120
 DB 130 IALNEADYVRWMLQADTMAAYQAVADAATVAVSTQPPAPRPGG-----DAAD 180
 QY 121 T-----AAANOLMNVNPOALKQLAOP 141
 DB 181 TRLDVLSSIGQLIRDI---LDFIANP 203

RESULT 8

SRA_MYCLE STANDARD; PRT; 408 AA.
 ID SRA_MYCLE STANDARD; PRT; 408 AA.
 AC Q07297;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine-rich antigen (25k) (45 kDa protein).
 GN SRA OR ML0411 OR MUC1383.14.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93239328; PubMed=8478104;
 RA Vega-Lopez F., Brooks L.A., Dockrell H.M., de Smet K.A.L.,
 RA Thompson J.K., Hussain R., Stoker N.G.;
 RT "Sequence and immunological characterization of a serine-rich antigen
 from Mycobacterium leprae";
 RL Infect. Immun. 61:2145-2153(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95020554; PubMed=7934845;
 RA Rinke de Wit T.F., Clark-Curtiss J.E., Abebe F., Kolkon A.H.J.,
 RA Jonson A.A.M., Thole J.E.R.;
 RT "A Mycobacterium leprae-specific gene encoding an immunologically
 recognized 45 kDa protein";
 RL Mol. Microbiol. 10:829-838(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RA "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00015; AAC43220.1; -
 CC EMBL; X68431; CAA48480.1; -
 CC EMBL; Z21952; CAA79350.1; -
 CC EMBL; Z97179; CAB09398.1; -

DR EMBL; AL583918; CAC29919.1; -
 DR PIR; C86960; C86960.
 DR PIR; S33522; S33522.
 DR PIR; S39872; S39872.
 DR LepToma; ML0411; -
 DR InterPro; IPR00030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Antigen; Repeat; Complete proteome.
 FT DOMAIN 182 196
 FT STRAIN=H37RV;
 FT MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogan A., Mclean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

RESULT 9

YU18 MYCTU STANDARD; PRT; 434 AA.
 ID YU18 MYCTU STANDARD; PRT; 434 AA.
 AC P31500; O53265;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv3018C/Mt3098/Mt3101.
 GN Rv3018C OR Mt3098/Mt3101 OR MTV012.32C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogan A., Mclean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

```

RC STRAIN=CDC 1551 / Oshkosh;
RR MEDLINE=22206494; PubMed=12218036;
RX Fletschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RRA Kolenay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RRA Biswal W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RRT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE OF 160-374 FROM N.A.
RR STRAIN=Isolate 50410;
RC Patki A.H., Dale J.W.;
RA Submitted (APR-1991) to the EMBL/GenBank/DDBJ databases.
RRL -! SIMILARITY: Belongs to the mycobacterial PPE family.
CC CC -! CAUTION: In strain Oshkosh the gene for this protein is
CC interrupted in position 307 by an IS6110 element.
CC -! CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
CC reductase.
CC -! CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 294; 337 and 355.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensed@isb-sib.ch).
CC -----
DR EMBL; AL021287; CAA16103.1; -.
DR EMBL; AS007129; AAK47427.1; ALT_SEQ.
DR EMBL; AS007129; AAK47430.1; ALT_SEQ.
DR EMBL; X59271; CAA41961.1; ALT_FRAME.
DR FIR; E70857; E70857.
DR TIGR; MT3096; -.
DR TIGR; MT3101; -.
DR Tuberculist; RV3018c; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE_1.
DR HypoNetical protein; Complete proteome.
SW SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;
SQ
Query Match 12.1%; Score 158.5; DB 1; Length 434;
Best Local Similarity 23.4%; Pred. No. 0.00053;
Matches 67; Conservative 34; Mismatches 130; Indels 55; Gaps 7;
QY 1 VAWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMLLIATNLLGNTTPA 60
DB ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 76 VAWLVQASADSAAGGEHEAAAAAGYGCALAEPTLPDLAAHHTHAVLVATNFPGINTIP 135
QY 61 IAVNEAYEGBMWAQDAAMFGYAAATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
DB ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 136 IALNEADYVPMVQAATWSAYEAVVGAAVLVAHPHGPAFVIVKPG-----ANEASN 187
QY 121 TAAAN-----QLMNVFOALKQLAQPTGGTTTSSKLGGMLWKTVSFH---RSPIS 166
DB ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 AVAAATITPPPWEHIEVFLEETFAAYDQYLSELLPA--VAWVWFOLFVDILGFNIIG 245
QY 167 NWYSMANNNMSMTNSGV-----MTNTLSMLKGPA 197
DB ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 246 FIITLASNAQLTEFAINASVAVGLLYATAGVIDIWEVIGNLFQGVPLLGPFLLGAL 305
QY 198 PAAAAAQAVQ-TAAQNGVRAMSSLGSSLGS-SGLGGGVAAANLGRAAS 241
DB ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 306 AAAVWPGVAGLAGVAGLAALPAVGAAAGAAPAALVGVSAVPSSGVGS 351
RESULT 10
YU21 MYCTU
ID YU21_MYCTU STANDARD; PRT; 435 AA.
AC O53268; O53269;

```

```

DT 16-OCT-2001 (Rel. 40, Created)
DT DT
DT 10-OCT-2003 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE HYPOCOTYL PPE-family protein RV3021C/RV3022C/MTR3106.
DE RV3021C/RV3022C OR MTR3106 OR MTRV012.35C/MTRV012.36C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxId=1773;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=H37Rv;
RC MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krohn A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishtai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol 184:5479-5490(2002).
RLL -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: Refl. sequence differs from that shown due to a
CC frameshift in position 82.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).

```

```

Query Match      12.1%; Score 158.5; DB 1; Length 434;
Best Local Similarity 23.4%; Pred. No. 0.00053;
Matches          67; Conservative 34; Mismatches 130; Indels 55; Gaps 7;

QY    1 VAWMSVTAGCAELTAQRVRAAAAYETAYGLTVPPVIAENRAELMLLIATNLLGONTPA 60
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     76 VAWIVQASDASAAGAHEHAAAAGYVCALAEMLTLPFLAANHITHAVLVATNFPGINTIP 135

QY    61 IAYNEAEYGEEMWAQDAAMAEGYAAATATATATLTPPEAPEMTSAGLLEBQAAAVEEASD 120
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     136 IALNEADYVRMWQOATVMSAYEAUVGAALVATPHTGPAPVIVKPG-----ANEASN 187

QY    121 TAAAN-----QLMNVQALKQLAQPTQGTFPSKLGGLWKTVSPH---RSPIS 166
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     188 AVAAAAITPPPHWHEIQQFLEETFAAYDQYLSELLSELP-VAVMWFQLFVDILGFNIIG 245

QY    167 NWYSMANNNHSMNTNSGVSY-----MTNTLSMLMKGFA 197
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     246 FIITLASNAQLLTFFAINASVAVGLLYATAGVIDVVEWVGNLFGVVPLGGPLLIAL 305

QY    198 PAAAAAQVQ-TAAQNQVRAMSSVLGSSLGS-SGLGGGVAAAMLGRAAS 241
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     306 AAAAVPGVAGLAGVAGLAALPAVGAAGAAPAALVGSVPVSGGVVS 351


RESULT 10
YU21_MYCTU
ID YU21_MYCTU STANDARD; PRT; 435 AA.
AC O53268; O53269;
```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ021287; CAA16106.1; ALT_FRAME.
DR EMBL: AJ021287; CAA16107.1; ALT_FRAME.
DR EMBL: A5007129; AAK47435.1; -.
DR TIGR: MT3106; -.
CC -----
DR Tuberculist; RV3021c; -.
DR Tuberculist; RV3022c; -.
DR InterPro; IPR00030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 299 299 G -> A (IN REF. 2).
FT CONFLICT 317 320 LAGV -> VTGL (IN REF. 2).
FT CONFLICT 326 326 L -> V (IN REF. 2).
SQ SEQUENCE 435 AA; 42876 MW; 3B157643EAA8484A CRC64;

Query Match : 12.1%; Score 158; DB 1; Length 435;
Best Local Similarity 21.9%; Pred. No. 0.00057;
Matches 67; Conservative 32; Mismatches 113; Indels 94; Gaps 7;

QY 1 VAWMSVYAGQAEILTAQVRVAAAAYETAYGLVPPPPVIAENRAELMILATNLLGNTFA 60
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 76 VAWLVQASADSAAGAEHEAAAGYVCAAEPMPTLPDLAANHUTHAVLVATNFFGINT:P 135
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 61 IAVNEAEYGVMAQDAADNMGVAA-----ATATATAT 92
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 136 IALNEADYVYRMVQAAATVMSAVEAVVGAALVATPHGTGPAPVIVKPGANEASNVAAATIT 195
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 93 LLPRFEE-----APENTSAGLLEQAAVEASDTAAANQLMNVPQALKQLAQPTQG 144
DB 196 PPFGLAKFLFLEMAQAFTFVGLIMKSAEAMAVGFVELITGLVNFEP----- 243
QY 145 TTPSSKLGGLKWTSPHRSPTSNMVM-----ANNHMSMTNSGV 183
DB 244 -----WLV-----LTGMIDMFFATVGFALGVFVLPLEFAVLELAILSIGW 286
QY 184 SMTNTLSS-----MLKGFAPAAAAQVQTAAGVGRVMSLSSGLGS-SGLGGGVAA 235
DB 287 IISNIFGAIPVLGFLGALAAAVVPGVAGLVAGLAALPAVGAAGAFAALVGSVAPV 346
QY 236 LGRAAS 241
DB 347 SGGVVS 352

RESULT 11

YY29 MYCTU STANDARD; PRT; 178 AA.
AC O06246;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3429/MT3533.
GN RV3429 OR MT3533 OR MTCY77.01.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
RN [1]
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z95389; CAB08678.1; --
CC EMBL; AE007158; AAK47873.1; --
CC PIR; C70975; C70975.
CC TIGR; MT3533; --
CC TubercuList; RV3429; --

DR InterPro: IPR00030; Microbac_PPE.
DR Pfam: PF0823; PPE; 1. Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 178 AA; 19811 MW; 8BE1FC025ABFBEA6 CRC64;
Query Match 11.9%; Score 156; DB 1; Length 178;
Best Local Similarity 37.1%; Pred. No. 0.00029;
Matches 39; Conservative 16; Mismatches 50; Indels 0; Gaps 0;
3 WMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTAIA 62
73 WLTKHRSQILEFAYVIDFLAVYEETRKVKVPPPTIANNREEVHRLIASNVAGVNTAIA 132
63 VNEAEYGEWMAQDAAMFGYAAATATATATATATATATATATATATATATATATATATAT 107
133 GLDAQYQYQYRAQNTIAVNDYQSTARFILAFLPRWQEPPIYGGGG 177

RESULT 12

YY25 MYCTU STANDARD; PRT; 176 AA.
AC Q50703;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV3425.
GN RV3425 OR MTCY78.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
RN [1]
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z95389; CAB08678.1; --
CC EMBL; AE007158; AAK47873.1; --
CC PIR; C70975; C70975.
CC TIGR; MT3533; --
CC TubercuList; RV3429; --

Db 133 DLDAQDQYRARNVAVNAYVWTRALSOLDLFWREPPQYRCG 176

RESULT 13

BUN2_DROME STANDARD; PRT; 1211 AA.

AC Q24523; Q9VK78; Q9VK79.

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE 10-OCT-2003 (Rel. 42, Last annotation update)

DE Bunched protein, class 2/class 3 isoforms (Shortsighted protein).

GN BUN OR SHS OR CG5461.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN NCBI_TaxID=7227;

RP SEQUENCE FROM N.A. (ISOFORM CLASS 2), SUBCELLULAR LOCATION, AND FUNCTION

RC TISSUE=Eye-antennal disk;

RX MEDLINE=96038094; PubMed=7555710;

RA Treisman J.E., Lai Z.-C., Rubin G.M.;

RT "Shortsighted acts in the decapentaplegic pathway in Drosophila eye development and has homology to a mouse TGF-beta-responsive gene.";

RL Development 121:2835-2845(1995).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS CLASS 2 AND CLASS 3).

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA April J.F., Agbayani A., An H.-J., Andrews-Franckoch C., Baldwin D.,

RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M., P.L.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furi V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas P., Testor C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

CC -!- FUNCTION: Probable transcription factor required for peripheral nervous system morphogenesis, eye development and oogenesis. May be required for the transmission of the app signal and for a morphogenetic movement of the medulla in the brain that reorients

the second optic lobe relative to the first. Plays a role in determining proper dorsal cell fates leading to the formation of the dorsal appendages.

!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.

!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=3;

Comment=Experimental confirmation may be lacking for some isoforms;

Name=Class 2;

Isoid=024523-1; Sequence=Displayed;

Name=Class 1;

Isoid=024522-1; Sequence=External;

Name=Class 3;

Isoid=024523-2; Sequence=VSP_006670;

!- SIMILARITY: Belongs to the TSC-22/Dip/Bun family.

!- CAUTION: Ref.2 sequence differs from that shown due to erroneous gene model prediction.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L42512; AAC41608.1; -

EMBL; AE003636; AAF53200.1; ALT SEQ.

EMBL; AE003636; AAF53201.1; ALT_SEQ.

PIR; T13804; T13804.

HSP; F80220; 1DIP.

FlyBase; Fgn00C10460; bun.

GO; GO:0007422; P:peripheral nervous system development; NAS.

InterPro; IPR000580; TSC-22_Dip_Bun.

Pfam; PF01166; TSC22; 1.

ProDom; PD007152; TSC-22_Dip_Bun; 1.

PROSITE; PS01289; TSC22_1.

Transcription regulation; Nuclear protein; Alternative splicing.

DOMAIN 15 31 POLY-GLN.

DOMAIN 76 86 POLY-GLN.

DOMAIN 97 102 POLY-SER.

DOMAIN 237 241 POLY-SER.

DOMAIN 249 254 POLY-SER.

DOMAIN 261 265 POLY-SER.

DOMAIN 306 321 POLY-HIS.

DOMAIN 322 328 POLY-GLN.

DOMAIN 348 356 POLY-GLY.

DOMAIN 607 619 POLY-GLN.

DOMAIN 661 668 POLY-GLN.

DOMAIN 743 746 POLY-ALA.

DOMAIN 759 765 POLY-GLN.

DOMAIN 817 822 POLY-GLN.

DOMAIN 832 838 POLY-ALA.

DOMAIN 884 891 POLY-ALA.

DOMAIN 927 947 GLN-RICH.

DOMAIN 1001 1005 POLY-ALA.

DOMAIN 1011 1014 POLY-ALA.

DOMAIN 1069 1090 LEUCINE-ZIPPER.

DOMAIN 1194 1201 POLY-ALA.

VARSPPLIC 1 109 Missing (in isoform Class 3).

FT CONFLICT 10 10 /FTIQ=VSP_006670.

FT CONFLICT 328 328 MISSING (IN REF. 1).

FT CONFLICT 801 801 Q -> QQQ (IN REF. 1).

FT CONFLICT 1189 1195 QVTSAA -> TS (IN REF. 2).

SEQ SEQUENCE 1211 AA; 125299 MW; 9925A9159A7051B0 CRC64;

Query Match 8.8%; Score 114.5; DB 1; Length 1211;

Best Local Similarity 25.9%; Pred. No. 1.8;

Matches 66; Conservative 28; Mismatches 80; Indels 81; Gaps 13;

QY 14 TAAQVRVAAAYETAYGLTVPPVTAENPAELMILLIATNLLCQNTFAIR---VNEAEYGE 70

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 19.156 Seconds
(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796C-26

Perfect score: 1306

Sequence: 1 VAMSVTAGQAEITAAQVRV.....YGHDDGGKYXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1187	90.9	391	16	O05298
2	1158.5	88.7	390	16	Q7U0E9
3	1122.5	85.9	396	16	Q7U071
4	1008.5	77.2	393	16	Q8VIZ3
5	949.5	72.7	393	16	Q06341
6	949.5	72.7	393	16	Q7TW55
7	463	35.5	399	16	Q7TW55
8	460	35.2	403	16	O53956
9	441	33.8	423	16	O53950
10	441	33.8	423	16	Q7T2I4
11	432.5	33.1	393	16	Q7T2I3
12	432.5	33.1	411	16	O53939
13	424.5	32.5	421	16	Q9Z5K0
14	422	32.3	409	16	O53957
15	420	32.2	409	16	Q7T2H7
16	420	32.2	410	2	Q99Q11

17	403	30.9	413	16	O06386	O06386 mycobacteri
18	402	30.8	391	16	P96362	P96362 mycobacteri
19	402	30.8	391	16	Q7U075	Q7U075 mycobacteri
20	390	29.9	394	16	Q7TXX5	Q7TXX5 mycobacteri
21	389	29.8	462	16	O33310	O33310 mycobacteri
22	387.5	29.7	694	16	Q8VJW0	Q8VJW0 mycobacteri
23	386.5	29.6	468	16	O53958	O53958 mycobacteri
24	379	28.0	380	16	P95190	P95190 mycobacteri
25	379	28.0	380	16	Q7TX66	Q7TX66 mycobacteri
26	375	28.7	385	16	O33204	O33204 mycobacteri
27	375	28.7	385	16	Q8VJZ0	Q8VJZ0 mycobacteri
28	375	28.7	385	16	Q7TZR7	Q7TZR7 mycobacteri
29	361	27.6	381	16	Q7TX67	Q7TX67 mycobacteri
30	355	27.2	382	16	Q7TXX3	Q7TXX3 mycobacteri
31	355	27.2	397	2	Q9AGF0	Q9AGF0 mycobacteri
32	352	27.0	402	16	O33312	O33312 mycobacteri
33	348.5	26.7	365	16	O86373	O86373 mycobacteri
34	347.5	26.6	405	16	Q8VJW5	Q8VJW5 mycobacteri
35	347	26.6	364	16	Q7TJZ5	Q7TJZ5 mycobacteri
36	346.5	26.5	394	16	O33205	O33205 mycobacteri
37	346.5	26.5	394	16	Q7TZR6	Q7TZR6 mycobacteri
38	346	26.5	350	16	Q7TJZ2	Q7TJZ2 mycobacteri
39	346	26.5	363	16	O53940	O53940 mycobacteri
40	339.5	26.0	443	16	Q8VKU9	Q8VKU9 mycobacteri
41	333.5	25.5	443	16	Q7U242	Q7U242 mycobacteri
42	332.5	25.2	187	16	Q7TZH6	Q7TZH6 mycobacteri
43	329	25.2	391	16	O05798	O05798 mycobacteri
44	328	25.2	391	16	Q7TX76	Q7TX76 mycobacteri
45	328	25.1	406	16	P71869	P71869 mycobacteri

ALIGNMENTS

RESULT 1

O05298	PRELIMINARY;	PRT;	391 AA.
ID	O05298		
AC	O05298;		
DT	01-JUL-1997 (TrEMBLrel. 04, Created)		
DT	01-JUL-1997 (TrEMBLrel. 04, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Hypothetical protein (PPE family protein).		
GN	RV1196 OR MTC1364.08 OR MT1234.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Cornebacteriaceae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37Rv;		
RX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,		
RA	Badcock K., Braham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,		
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,		
RA	Rutter S., Seeger K., Skelton S., Squares R.,		
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;		
RA	"Deciphering the Biology of Mycobacterium tuberculosis from the		
RT	complete genome sequence."		
RL	Nature 393:537-544(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CDC 1551 / Oshkosh;		
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,		
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,		
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,		
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,		
RA	Bisler W.;		
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and		
RT	laboratory strains."		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		


```
DE PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RC STRAIN=FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AEO07161; AAK47941.1; -.
DR TIGR; MT3582; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE_1; 1.
SQ SEQUENCE 393 AA; 33688 MW; 86F0B67798855511 CRC64;

Query Match 77.2%; Score 1008.5; DB 16; Length 393;
Best Local Similarity 84.6%; Pred. No. 6.6e-61;
Matches 208; Conservative 14; Mismatches 19; Indels 5; Gaps 2;

QY 1 VAMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
DB 74 VAMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133

QY 61 IAYNEAEYGENWAQDAAMFGYAAATATATATATLPPFEAPEMTSAGLLEQAAVVEEASD 120
DB 134 IAYNEAEYGENWAQDAAMFGYAAATATATATLPPFEAPEMTSAGLLEQAAVVEEASD 193

QY 121 TAANQLMNNVPQALQKLAQPTGTTSSKLGGLWKTVPSPISNMVSMANNHSMWN 180
DB 194 TAANQLMNNVPQALQKLAQPTGTTSSKLGGLWKTVPSPISNMVSMANNHSMWN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAVTAQNGVRAMSS---LGSSLSGSGGGGVAANL 236
DB 254 SGVSMNTLSSMLKGFAPAAAQAVTAQNGVRAMSS---LGSSLSGSGGGGVAANL 312

QY 237 GRAASV 242
DB 313 GRAASV 318

RESULT 5
Q06341 PRELIMINARY; PRT; 393 AA.
AC O06341;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR WTC13E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RC STRAIN=H37RV;
RC MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
```

```
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; Z95390; CAB08702.1; -.
DR FIR; C70568; C70568.
DR Tuberculist; RV3478; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 72.7%; Score 949.5; DB 16; Length 393;
Best Local Similarity 79.7%; Pred. No. 7e-57;
Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;

QY 1 VAMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
DB 74 VAMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133

QY 61 IAYNEAEYGENWAQDAAMFGYAAATATATATLPPFEAPEMTSAGLLEQAAVVEEASD 120
DB 134 IAYNEAEYGENWAQDAAMFGYAAATATATATLPPFEAPEMTSAGLLEQAAVVEEASD 193

QY 121 TAANQLMNNVPQALQKLAQPTGTTSSKLGGLWKTVPSPISNMVSMANNHSMWN 180
DB 194 TAANQLMNNVPQALQKLAQPTGTTSSKLGGLWKTVPSPISNMVSMANNHSMWN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAVTAQNGVRAMSS---LGSSLSGSGGGGVAANL 236
DB 254 SGVSMNTLSSMLKGFAPAAAQAVTAQNGVRAMSS---LGSSLSGSGGGGVAANL 312

QY 237 GRAASV 242
DB 313 GRAASV 318

RESULT 6
Q07WF5 PRELIMINARY; PRT; 393 AA.
AC Q07WF5;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE PE family protein.
GN PPE60 OR ME3505.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1765;
RN [1]
RC STRAIN=AF2122/97;
RC MEDLINE=22709107; PubMed=12789972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Iacox C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248346; CAD95692.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 72.7%; Score 949.5; DB 16; Length 393;
Best Local Similarity 79.7%; Pred. No. 7e-57;
Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;

QY 1 VAMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
```

Db 74 VAWMSVTAGQQLTAQVRVAAAAYETAYRTVTPPPVIAENRTEMTLTATNLLGQNTPA 133
 QY 61 IAVNEAEYGEWMAQDAAMFYAAATATATATATATLTPFEAPEMTSAGGLLEQAAVEASD 120
 Db 134 IEANQAAYSQWQDAEMYGVAATAATATATATATALLPFDAPLITNPGGLLEQAAVEAID 193
 QY 121 TAAANQNNVPOALKQALQAOPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMTN 180
 Db 194 TAAANQNNVPOALKQALQAOPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMWG 253
 QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMGS----LGSSGLSSGLGGGVAANL 236
 Db 254 TGVSMNTLHSMKGLAP-AAAQAVETAAENGVMAMSSLSGLSSGLSSGLGGGVAANL 312
 QY 237 GRAASV 242
 Db 313 GRAASV 318

RESULT 7

Q7TZH8 PRELIMINARY; PRT; 399 AA.
 AC Q7TZH8;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DE PPE family protein.
 GN PP31 OR M1836.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12789972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248340; CAD94539.1; -
 KW Complete proteome.
 SQ SEQUENCE 399 AA; 38840 MW; 1A0F4377318E74F2 CRC64;

Query Match 35.5%; Score 463; DB 16; Length 399;
 Best Local Similarity 43.7%; Pred. No. 1e-23;
 Matches 118; Conservative 36; Mismatches 72; Indels 44; Gaps 9;

QY 1 VAWMSVTAGQELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
 Db 73 VAWMSVTAVRAEQAGAQAEAAAAYEAAFAATVPPVIEANRAQMALIATNVLGQVAP 132
 QY 61 IAVNEAEYGEWMAQDAAMFYAAATATATATLTPFEAPEMTSAGGLLEQAAVEASD 120
 Db 133 IATEAQAEYGEWMAQDAAMFYAAATATATLTPFEAPEMTSAGGLLEQAAVEASD 191
 QY 121 TAAA-----NQLMNNVPQALKQALQAOPTQGT-----PSSKLG-----GLWKTVPSPH 161
 Db 192 ASAGAQQTTLSQLIAAIPSVLQGLSSSTAATSASGPSGLLGILGSSGLDLDKLLWLLDPN 251
 QY 162 RSPISNMVMANNHMSMTNSGVSM-TNTLS---SMLKGFAPAAAQAVQTAQAQNGVRAMS 217
 Db 252 -----SNFWNTIASSGLFLPSNTIAPFLGLGGVAAAADAAAGDVLGEATSG----- 296
 QY 218 SLGSSL-----GSSGLGGGVAANLGRAASV 242
 Db 297 GLGGALVAPLGSAGGLGGTVAAGLGNATV 326

RESULT 8

OS3956 PRELIMINARY; PRT; 403 AA.
 AC OS3956;
 DT 01-JUN-1998 (TREMELrel. 06, Created)
 DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE PPE-family protein.
 GN RV1807 OR MT1856 OR MTV049.29.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.B. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL022021; CAAL7728.1; -
 DR EMBL; AE007044; AAK46128.1; -
 DR F1; H70931; H70931.
 DR TIGR; MT1856; -
 DR TubercuList; RV1807; -
 DR InterPro; IPR000300; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome.
 FT CONFLICT 227 227 F -> S (IN REF. 2).
 FT CONFLICT 238 238 V -> L (IN REF. 2).
 SQ SEQUENCE 403 AA; 39243 MW; DCE18680FD15CBPE CRC64;

Query Match 35.2%; Score 450; DB 16; Length 403;
 Best Local Similarity 43.3%; Pred. No. 1.7e-23;
 Matches 117; Conservative 36; Mismatches 73; Indels 44; Gaps 8;

QY 1 VAWMSVTAGQELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
 Db 77 VAWMSVTAVRAEQAGAQAEAAAAYEAAFAATVPPVIEANRAQMALIATNVLGQVAP 136
 QY 61 IAVNEAEYGEWMAQDAAMFYAAATATATLTPFEAPEMTSAGGLLEQAAVEASD 120
 Db 137 IATEAQAEYGEWMAQDAAMFYAAATATLTPFEAPEMTSAGGLLEQAAVEASD 195
 QY 121 TAAA-----NQLMNNVPQALKQALQAOPTQGT-----TPSSKLGGLWKTVPSPH 161
 Db 196 ASAGAQQTTLSQLIAAIPSVLQGLSSSTAATFASGPSGLLGIVGSSGLDLDKLLWLLDPN 255
 QY 162 RSPISNMVMANNHMSMTNSGVSM-TNTLS---SMLKGFAPAAAQAVQTAQAQNGVRAMS 217
 Db 256 -----SNFWNTIASSGLFLPSNTIAPFLGLGGVAAAADAAAGDVLGEATSG----- 300
 QY 218 SLGSSL-----GSSGLGGGVAANLGRAASV 242
 Db 301 GLGGALVAPLGSAGGLGGTVAAGLGNATV 330

Db 284 KIGGKPTGALAPLAEFALHTPIILGSEGLGGSVSAGIGRAGLV 327

RESULT 9

OS3950 PRELIMINARY; PRT; 423 AA.
 AC OS3950;
 DT 01-JUN-1998 (TREMELrel. 06, Created)
 DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE PPE family protein.
 GN RV1801 OR MTU049.23 OR MT1850.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Unayam L.A., Emdolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR ENBL; AL020201; CAAL7722.1; -.
 DR ENBL; AE007044; AAK46122.1; -.
 DR PIR; B70931; B70931.
 DR TIGR; MT1850; -.
 DR TubercuList; RV1801; -.
 DR InterPro; IPR000030; Microbac_ppe.
 DR Pfam; PF00823; ppe; 1.
 KW Complete proteome.
 SQ SEQUENCE 423 AA; 41477 MW; 26E52CC271FBBF57 CRC64;

Query Match 33.8%; Score 441; DB 16; Length 423;
 Best Local Similarity 41.2%; Pred. No. 3.5e-22;
 Matches 117; Conservative 30; Mismatches 67; Indels 70; Gaps 9;
 QY 1 VAWMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA 60
 Db 72 VAWMSATVABELAGQARLAIAYEAFAATVPPVIAANRAQLMVLITNIFGQNTPA 131
 QY 61 IAVNEAEGEMWAQDAAMFGYAAATATATATATLPPFEAPEMTSAGLLEQAAAVEASD 120
 Db 132 IMVTEAQYMENWAQDAAMFGYAGSSATA-SRMTAFTEPPQTTNHGOLGAQSSAVAQTA 190
 QY 121 TAAAN-----OLMNVVPOALKOLAQPT-----QGTFP-----SSKLGGLWKTVP 160
 Db 191 TAAGNLOSAPFOLLAVPRALQGLALPTASQSASATPQWTDLGNLSTFLGG--AVTGP 248
 QY 161 HRSPISNNVSMANNHMTNSGVSMTNTLSMLKGFAPAAAAQAVQTAQNGVRAMSSLG 220
 Db 249 YTFP-----GVLPSPGVPLLIGISVL-----VTQNGQGVSAALLG 283
 QY 221 S-----SIGSSGLGGG-VANLGRAASV 242

RESULT 11

OS3950 PRELIMINARY; PRT; 393 AA.
 AC OS3950;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE26 OR MB1817.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,

Db 284 KIGGKPTGALAPLAEFALHTPIILGSEGLGGSVSAGIGRAGLV 327
 RESULT 10
 Q7TZI4
 ID Q7TZI4 PRELIMINARY; PRT; 423 AA.
 AC Q7TZI4;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE PPE family protein.
 GN PPE29 OR MB1829.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR ENBL; BX248340; CAD94532.1; -.
 KW Complete proteome.
 SQ SEQUENCE 423 AA; 41477 MW; 26E52CC271FBBF57 CRC64;

Query Match 33.8%; Score 441; DB 16; Length 423;
 Best Local Similarity 41.2%; Pred. No. 3.5e-22;
 Matches 117; Conservative 30; Mismatches 67; Indels 70; Gaps 9;
 QY 1 VAWMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA 60
 Db 72 VAWMSATVABELAGQARLAIAYEAFAATVPPVIAANRAQLMVLITNIFGQNTPA 131
 QY 61 IAVNEAEGEMWAQDAAMFGYAAATATATATATLPPFEAPEMTSAGLLEQAAAVEASD 120
 Db 132 IMVTEAQYMENWAQDAAMFGYAGSSATA-SRMTAFTEPPQTTNHGOLGAQSSAVAQTA 190
 QY 121 TAAAN-----OLMNVVPOALKOLAQPT-----QGTFP-----SSKLGGLWKTVP 160
 Db 191 TAAGNLOSAPFOLLAVPRALQGLALPTASQSASATPQWTDLGNLSTFLGG--AVTGP 248
 QY 161 HRSPISNNVSMANNHMTNSGVSMTNTLSMLKGFAPAAAAQAVQTAQNGVRAMSSLG 220
 Db 249 YTFP-----GVLPSPGVPLLIGISVL-----VTQNGQGVSAALLG 283
 QY 221 S-----SIGSSGLGGG-VANLGRAASV 242

RESULT 11

OS3950 PRELIMINARY; PRT; 393 AA.
 AC OS3950;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE26 OR MB1817.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,

RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248340; CAD94520.1; -.
 KW Complete proteome.
 SQ SEQUENCE 393 AA; 38572 MW; 41944B6E547A8AB0 CRC64;

Query Match 33.1%; Score 432.5; DB 16; Length 393;
 Best Local Similarity 41.0%; Pred. No. 1.2e-21;
 Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
 QY 1 VAWMSVTAQAGELTAQAQVRAAAAYETAYGLTVPPVIAENRAELMILIAINLLQNTTPA 60
 Db 73 VAWMSAAAQAQAEQAATQARAAAAAFAAFAATVPPPLIAANRASLMQLISTNVFQNTSA 132
 QY 61 IAVNEAEYGEWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAVEEASD 120
 Db 133 IAAEAQYGEWAQDAAMFYAGSSASASA-VTFSTPPQIANPTAQTQAAAVATAAG 191
 QY 121 TAAA--NQLMNVPOALKOLAQTGTTTSSKLGGLMK-----TVSPH 161
 Db 192 TAQSTLTEMITGLPNALQSLTSPLLQSS-NGPLSLWQLFGTFNFTSISALLTDLPY 250
 QY 162 RPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAQNGVRAMSSLS 221
 Db 251 ASFFYNTGLPYFSGMGNFTQAATL-GLIGSAAPAAVA-----AAGDAAGLPLGG 304
 QY 222 SLGSSGLGGVAANLGRAASV 242
 Db 305 MLG-----GGPVAAGLGNRAASV 321

RESULT 12

OS3939 ID O53939 PRELIMINARY; PRT; 411 AA.
 AC O53939;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE PPE-family protein.
 GN RV1789 OR MT1838 OR MT049.11.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaita F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL022021; CAAL17711.1; ALT_INIT.
 DR EMBL; AE007043; AAK46108.1; -.
 DR PIR; G70929; G70929.
 DR TIGR; MT1838; -.
 DR Tuberculist; Rv1789; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome.
 SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 33.1%; Score 432.5; DB 16; Length 411;
 Best Local Similarity 41.0%; Pred. No. 1.3e-21;
 Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
 QY 1 VAWMSVTAQAGELTAQAQVRAAAAYETAYGLTVPPVIAENRAELMILIAINLLQNTTPA 60
 Db 91 VAWMSAAAQAQAEQAATQARAAAAAFAAFAATVPPPLIAANRASLMQLISTNVFQNTSA 150
 QY 61 IAVNEAEYGEWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAVEEASD 120
 Db 151 IAAEAQYGEWAQDAAMFYAGSSASASA-VTFSTPPQIANPTAQTQAAAVATAAG 209
 QY 121 TAAA--NQLMNVPOALKOLAQTGTTTSSKLGGLMK-----TVSPH 161
 Db 210 TAQSTLTEMITGLPNALQSLTSPLLQSS-NGPLSLWQLFGTFNFTSISALLTDLPY 268
 QY 162 RPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAQNGVRAMSSLS 221
 Db 269 ASFFYNTGLPYFSGMGNFTQAATL-GLIGSAAPAAVA-----AAGDAAGLPLGG 322
 QY 222 SLGSSGLGGVAANLGRAASV 242
 Db 323 MLG-----GGPVAAGLGNRAASV 339

RESULT 13

Q925K0 ID Q925K0 PRELIMINARY; PRT; 421 AA.
 AC Q925K0;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative PPE protein (PPE-family protein).
 GN ML1182 OR MLCB1701.08C.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 DR EMBL; AL049191; CAB39148.1; -.
 DR EMBL; AL583921; CAB31563.1; -.
 DR PIR; H87056; H87056.
 DR Leproma; ML1182; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome.
 SQ SEQUENCE 421 AA; 43119 MW; 680FD5FDCB65B4A7 CRC64;

Query Match 32.5%; Score 424.5; DB 16; Length 421;

[illegible]

```

RESULT 14
OS03957
ID 053957 PRELIMINARY; PRT; 409 AA.
AC 053957;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1808 OR MT1856.1 OR MT049.30.
GS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
SEQUENCE FROM N.A.
RN
RP
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teklaia E.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
SEQUENCE FROM N.A.
RN
RP
RC STRAIN=CDC 1551 / OshKosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.J., Nelson W.C., Unayam L.A., Emdolaeva M., Salzberg S.D.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AJ022021; CAA17729.1; -.
DR ENBL; AE007044; AAK46129.1; ALT_INIT.
DR PIR; A70932; A70932.
DR TIGR; MT1856.1; -.
DR Tuberculist; Rv1808; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Complete proteome.
SQ
SEQUENCE 409 AA; 39917 MW; 1E15202BACF36379 CRC64;

```

Query Match	32.3%	Score	422	DB	16	Length	409
Best Local Similarity	42.2%	Pred.	No.	6.6e-21			
Matches	108	Conservative	34	Mismatches	92	Indels	22
Gaps	8						
QY	1	VAMSVTACCAELTAQOEVAAAYETAYGLTVPPVIAEAEMLIATNLGQNTPA	60				
DB	73	VAMSVTACCAQAGAKIAGVETAFAAVPPVIEANRALLMSLVATNIFGQNTPA	132				
QY	61	IAVNEAAYGEMWAQDAAMAFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEASD	120				
DB	133	IAATEAHYAEMWAQDAAMAFGYAGSSATA-SQAPFSEPPQTTPSATAQSAVVAQAAG	191				
QY	121	TAAA-----NQLMNVDPQALKQLAOPTCGGTTPSSKLGGLWKTVPSPHSPISNMYSMAN	173				
DB	192	AAASDDITACLQLSLLEPSTICLSLA--TATATASAG--WDTV---LQSITITILANLT	244				
QY	174	NHMSMTNSGV-----MTNTLSMLKGRAPAAAQAQVCTAAQNGVRAMSL-CSSLSGS-S	256				
DB	245	GYSIIGLGAICGGWLTTCGILGLAQNAPGVALLGPKAAAGALSPLAPLRGGYIGDIT	304				
QY	227	GLGGGVAANLGRAASV 242					
DB	305	PLGGGATGGIARAIV 320					

RESULT 15

Q7TZH7

ID

Q7TZH7

PRELIMINARY;

PRT;

409 AA.

AC

Q7TZH7;

01-OCT-2003 (TrEMBLrel. 25, Created)

DT

01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE

PPE family protein.

GN

pPE32 OR MB1837.

OS

Mycobacterium bovis.

OC

Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;

OC

Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OX

NCBI_TaxID=1765;

[1]

SEQUENCE FROM N.A.

RP

STRAIN=AF2122/97;

RC

MEDLINE=22709107; PubMed=12788972;

RA

Garnier T., Eighmeier K., Camus J.-C., Medina N., Mansoor H.,

RA

Pryor M., Duthoy S., Grondin S., Iacox C., Monsemp C., Simon S.,

RA

Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,

RA

Parthill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;

RT

"The complete genome sequence of Mycobacterium bovis."

RL

Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).

DR

EMBL; BX248340; CAD94540.1; --.

KW

Complete proteome.

SQ

SEQUENCE 409 AA; 39931 MW; D57892628B131A9E CRC64;

[illegible]

Db : |||| : || |
304 TPLGGGATGGIARAIYV 320

Search completed: June 30, 2004, 16:52:53
Job time : 20.156 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 22.1761 seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796c-12
Perfect score: 3686
Sequence: 1 MHHHHHTAASDNFQSQGG.....SGGPVWNGVGWGMNTAAS 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3686	100.0	729	4	US-09-223-040-2
2	3686	100.0	729	4	US-09-287-849-2
3	2941	79.8	596	4	US-09-287-849-26
4	1967.5	53.4	600	4	US-09-287-849-22
5	1944	52.7	391	3	US-08-818-112-107
6	1944	52.7	391	4	US-08-818-111-102
7	1944	52.7	391	4	US-09-056-556-107
8	1944	52.7	391	4	US-09-072-596-102
9	1944	52.7	391	4	US-09-072-967-107
10	1525.5	44.8	396	3	US-08-818-112-111
11	1525.5	44.8	396	4	US-08-818-111-106
12	1525.5	44.8	396	4	US-09-056-556-111
13	1525.5	44.8	396	4	US-09-072-596-106
14	1525.5	44.8	396	4	US-09-072-967-111
15	1486.5	40.3	359	3	US-08-818-112-109
16	1486.5	40.3	359	4	US-08-818-111-104
17	1486.5	40.3	359	4	US-09-056-556-109
18	1486.5	40.3	359	4	US-09-072-596-104
19	1486.5	40.3	359	4	US-09-072-967-109
20	1184	32.1	358	4	US-09-287-849-8
21	1182	32.1	263	3	US-08-818-112-91
22	1182	32.1	263	4	US-08-818-111-92
23	1182	32.1	263	4	US-09-056-556-91
24	1182	32.1	263	4	US-09-072-596-92
25	1182	32.1	263	4	US-09-072-967-91
26	983	26.7	355	3	US-08-818-112-79
27	983	26.7	355	4	US-08-818-111-80

ALIGNMENTS

RESULT 1
US-09-223-040-2
; Sequence 2, Application US/09223040 ✓
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-223-040-2

Query Match	100.0%;	Score 3686;	DB 4;	Length 729;
Best Local Similarity	100.0%;	Pred. No. 3.6e-261;		
Matches 729;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MHHHHHTAASDNFQSQGGGPAIPICGAWAIAAGQIRSGGSPVHIGTAFILGLGVVD	60	
Db	1	MHHHHHTAASDNFQSQGGGPAIPICGAWAIAAGQIRSGGSPVHIGTAFILGLGVVD	60	
QY	61	NNGNGARVQVVGSAPAAASLGISTGVDITAVDGPINSATAMADALNGHHHPGDIIVTQ	120	
Db	61	NNGNGARVQVVGSAPAAASLGISTGVDITAVDGPINSATAMADALNGHHHPGDIIVTQ	120	
QY	121	TKSGGTRTGNVTLAEGPPAEFMDFGALPPINSGARMYAGPGSASLVAAAMWDSVASDL	180	
Db	121	TKSGGTRTGNVTLAEGPPAEFMDFGALPPINSGARMYAGPGSASLVAAAMWDSVASDL	180	
QY	181	PSASAFQSVVWGLTVGWSGSSAGLWVAASPVVAMSVTAAOAGELTAACVRAAAAYE	240	
Db	181	PSASAFQSVVWGLTVGWSGSSAGLWVAASPVVAMSVTAAOAGELTAACVRAAAAYE	240	
QY	241	TAYGLTVPFPVIAENRAELMILITNLGQNTPAIVNEAFYGEWQAQDAAMFGYAAAT	300	
Db	241	TAYGLTVPFPVIAENRAELMILITNLGQNTPAIVNEAFYGEWQAQDAAMFGYAAAT	300	
QY	301	ATATATLLPFEAEPMISAGLLECAAAVEASTAAANOLMNVPOALQLOAQPTQGT	360	

Db 301 ATATATLPPFEAPBMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGTT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOQAWAANAQVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOQAWAANAQVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLVPPRPYVMPHSPAAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLVPPRPYVMPHSPAAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
Db 601 GATDINAFSVSGGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 2

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2
Query Match 100.0%; Score 3686; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.6e-26;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFQISQGGQGFAPIGQAWAIAQIRSGGSPVTHIGPTAFILGLGVVD 60
|||||

Db 1 MHHHHHTAASDNFQISQGGQGFAPIGQAWAIAQIRSGGSPVTHIGPTAFILGLGVVD 60
QY 61 NNGNGARVORVVGSAAPASLSLGISTGDVITAVDGAIPNSATAMADALNGHHFGDDVISVTWQ 120
Db 61 NNGNGARVORVVGSAAPASLSLGISTGDVITAVDGAIPNSATAMADALNGHHFGDDVISVTWQ 120
QY 121 TKSQGTTRGNTVTLAEGPPAEFVWDFGALPPPEINSARMTYAGPGSASLVAAAQMWDSVASDL 180
Db 121 TKSQGTTRGNTVTLAEGPPAEFVWDFGALPPPEINSARMTYAGPGSASLVAAAQMWDSVASDL 180
QY 181 FSAASAFOSVVMGLTVGSGWIGSSAGLVAAASPYVAMSVTAGOAEITAAQVRRVAAAAYE 240
Db 181 FSAASAFOSVVMGLTVGSGWIGSSAGLVAAASPYVAMSVTAGOAEITAAQVRRVAAAAYE 240
QY 241 TAYGLTVPPPVIENRAELMILIAITNLLGQNTPAIENAEYGENWAOQAAAAAFYAAAT 300
Db 241 TAYGLTVPPPVIENRAELMILIAITNLLGQNTPAIENAEYGENWAOQAAAAAFYAAAT 300
QY 301 ATATATLPPFEAPBMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGTT 360
Db 301 ATATATLPPFEAPBMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGTT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOQAWAANAQVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOQAWAANAQVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLVPPRPYVMPHSPAAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLVPPRPYVMPHSPAAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
Db 601 GATDINAFSVSGGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 3

US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 79.8%; Score 2941; DB 4; Length 596;
Best Local Similarity 99.8%; Pred. No. 7.6e-207;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 142 MYDFGALPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201
DB 9 MYDFGALPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
QY 202 SSAGLMVAAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
DB 69 SSAGLMVAAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
QY 262 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 321
DB 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 322 LLEQAAAVEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 381
DB 189 LLEQAAAVEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
QY 382 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQTAQNGVRAMSSLSGSSG 441
DB 249 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQTAQNGVRAMSSLSGSSG 308
QY 442 LGGVVAANLGRAASVGSLSVPOWAAAANOAVTPAARALPLTSLTSAARPGQMLGGLPV 501
DB 309 LGGVVAANLGRAASVGSLSVPOWAAAANOAVTPAARALPLTSLTSAARPGQMLGGLPV 368
QY 502 GQVNVNINIKLGYNNAGAGTGVDPNGVLTNNHVIAGATDINAFSGSGQTVGVDDV 621
DB 429 GQVNVNINIKLGYNNAGAGTGVDPNGVLTNNHVIAGATDINAFSGSGQTVGVDDV 488
QY 622 GYDRTQDVAVLQRLGAGGLPSAAICGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQT 681
DB 489 GYDRTQDVAVLQRLGAGGLPSAAICGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQT 548
QY 682 VQASDLSLTGAETLNGLIQFDAAIQGDSGGPVVNGLGQVGMNTAAS 729
DB 549 VQASDLSLTGAETLNGLIQFDAAIQGDSGGPVVNGLGQVGMNTAAS 596

RESULT 4

US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US

; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 53.4%; Score 1967.5; DB 4; Length 600;
Best Local Similarity 72.0%; Pred. No. 1e-135;
Matches 431; Conservative 19; Mismatches 82; Indels 67; Gaps 10;
QY 142 MYDFGALPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201
DB 9 MYDFGALPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
QY 202 SSAGLMVAAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
DB 69 SSAGLMVAAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
QY 262 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 321
DB 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 322 LLEQAAAVEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 381
DB 189 LLEQAAAVEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
QY 382 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQTAQNGVRAMSSLSGSSG 441
DB 249 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQTAQNGVRAMSSLSGSSG 308
QY 442 LGGVVAANLGRAASVGSLSVPOWAAAANOAVTPAARALPLTSLTSAARPGQMLGGLPV 501
DB 309 LGGVVAANLGRAASVGSLSVPOWAAAANOAVTPAARALPLTSLTSAARPGQMLGGLPV 368
QY 502 GQVNVNINIKLGYNNAGAGTGVDPNGVLTNNHVIAGATDINAFSGSGQTVGVDDV 621
DB 429 GQVNVNINIKLGYNNAGAGTGVDPNGVLTNNHVIAGATDINAFSGSGQTVGVDDV 488
QY 622 GYDRTQDVAVLQRLGAGGLPSAAICGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQT 681
DB 489 GYDRTQDVAVLQRLGAGGLPSAAICGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQT 548
QY 682 VQASDLSLTGAETLNGLIQFDAAIQGDSGGPVVNGLGQVGMNTAAS 729
DB 549 VQASDLSLTGAETLNGLIQFDAAIQGDSGGPVVNGLGQVGMNTAAS 596

RESULT 5

US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedwick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 153
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/818,112
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C6
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 107:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-818-112-107

Query Match 52.7%; Score 1944; DB 3; Length 391;
 Best Local Similarity 99.7%; Pred. No. 3e-134;
 Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	142	MVDFGALPPPEINSGARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSMIG	201
Db	1	MVDFGALPPPEINSGARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSMIG	60
Qy	202	SSAGLMVAASPPVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI	261
Db	61	SSAGLMVAASPPVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI	120
Qy	262	LIATNLGQNTPAIAVNEAEYGEWMAQDAAAAFGYAAATATATATATLLPPEEAPEMTSAGG	321
Db	121	LIATNLGQNTPAIAVNEAEYGEWMAQDAAAAFGYAAATATATATLLPPEEAPEMTSAGG	180
Qy	322	LLEQAAVEEASDTAAANQLMNNVPAALQQLAQPTGTTTSSKLGGLWKTVPSPHSPISN	381
Db	181	LLEQAAVEEASDTAAANQLMNNVPAALQQLAQPTGTTTSSKLGGLWKTVPSPHSPISN	240
Qy	382	MVSMANNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG	441
Db	241	MVSMANNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG	300
Qy	442	LGGSVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV	501
Db	301	LGGSVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV	360
Qy	502	GQMGARAGGGLSGVLRVPRPYWPHSPAAG	532
Db	361	GQMGARAGGGLSGVLRVPRPYWPHSPAAG	391

RESULT 6
 US-08-818-111-102
 ; Sequence 102, Application US/08818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedwick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/818,111
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 102:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-818-111-102

Query Match 52.7%; Score 1944; DB 4; Length 391;
 Best Local Similarity 99.7%; Pred. No. 3e-134;
 Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	142	MVDFGALPPPEINSGARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSMIG	201
Db	1	MVDFGALPPPEINSGARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSMIG	60
Qy	202	SSAGLMVAASPPVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI	261
Db	61	SSAGLMVAASPPVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI	120
Qy	262	LIATNLGQNTPAIAVNEAEYGEWMAQDAAAAFGYAAATATATATLLPPEEAPEMTSAGG	321
Db	121	LIATNLGQNTPAIAVNEAEYGEWMAQDAAAAFGYAAATATATATLLPPEEAPEMTSAGG	180
Qy	322	LLEQAAVEEASDTAAANQLMNNVPAALQQLAQPTGTTTSSKLGGLWKTVPSPHSPISN	381
Db	181	LLEQAAVEEASDTAAANQLMNNVPAALQQLAQPTGTTTSSKLGGLWKTVPSPHSPISN	240
Qy	382	MVSMANNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG	441
Db	241	MVSMANNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSLGSSG	300
Qy	442	LGGSVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV	501
Db	301	LGGSVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV	360

Qy 502 GQMGARAGGGLSGVLRVPPRPYVMHSPAAG 532

Db 361 GQMGARAGGGLSGVLRVPPRPYVMHSPAAG 391

RESULT 7

RECORD. 1
US-09-056-556-107
; Sequence 107, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-107

```
Query Match      52.7%; Score 1944; DB 4; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134;
Matches 390: Conservative 0; Mismatches 1; Indels 0; Caps 0;
```

QY	142	MVDFGALPPIN	ARMYTAGP	GSASLVAA	QWDSVASDL	FSAA	SAFOSV	VWGLTV	GSWIG	201
DB	1	MVDFGALPPIN	ARMYTAGP	GSASLVAA	QWDSVASDL	FSAA	SAFOSV	VWGLTV	GSWIG	60
QY	202	SSAGLMAVAA	SPYVAM	SVTAGO	AELTAAQVR	VAAAAAYETAY	GLTVPP	PVIAEN	RAELMI	261
DB	61	SSAGLMAVAA	SPYVAM	SVTAGO	AELTAAQVR	VAAAAAYETAY	GLTVPP	PVIAEN	RAELMI	120
QY	262	LIATNLLGONT	PATAVNEAE	YGEWMAQD	AAAMFCYAAATATATAT	ATLTPP	FEAP	APMT	SAGG	321
DB	121	LIATNLLGONT	PATAVNEAE	YGEWMAQD	AAAMFCYAAATATATAT	ATLTPP	FEAP	APMT	SAGG	180
QY	322	LLEQAAAVEE	ASDTPAAANQ	LNNVPQALQ	LAQPTQGTTP	SSKLGGLWKT	VS	PHRS	PISN	381
DB	181	LLEQAAAVEE	ASDTPAAANQ	LNNVPQALQ	LAQPTQGTTP	SSKLGGLWKT	VS	PHRS	PISN	240
QY	382	MVSMANNH	MSNTSGV	SMNTTLL	SMWLKGFAPAA	ARQAVQTAAQNG	VRAMS	LSGSSL	GSSG	441
DB	241	MVSMANNH	MSNTSGV	SMNTTLL	SMWLKGFAPAA	ARQAVQTAAQNG	VRAMS	LSGSSL	GSSG	300
QY	442	IGGGVAAN	LGSAAS	VGSLSP	QVPAWAAANQAVT	PAARALPLT	SLTSA	AABERG	QOMGLGPV	501
DB	301	IGGGVAAN	LGSAAS	VGSLSP	QVPAWAAANQAVT	PAARALPLT	SLTSA	AABERG	QOMGLGPV	360

502 GQMGARAGGGLSGVLRVPPRPYVNPSPAAAG 532

361 GQMGARAGGGLSGVLRVPPRPYVNPSPAAAG 391

RESULT 8

US-09-072-596-102
 ; Sequence 102, Application US/09072596
 ; Patent No. 6458366
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Meto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,596
 ; FILING DATE: 05-MAY-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 102:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-072-596-102

Query Match	52.7%;	Score 1944;	DB 4;	Length 391;
Best Local Similarity	99.7%;	Pred. No. 3e-134;		
Matches 390;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps	0;			

142	QY	MYDFGALPPEINSARMYAGFGCSASLVAAQAQWDSVASDIFSAAAFQSVVWGLTWTGVSMTG	201
		1 MYDFGALPPEINSARMYAGFGCSASLVAAQAQWDSVASDIFSAAAFQSVVWGLTWTGVSMTG	60
202	QY	SSAGLMWAAAAPYVAMWSVTAGQAELTAAQVRVAAAAYETAVGLTVPPEVIAENRAELMI	261
		61 SSAGLMWAAAAPYVAMWSVTAGQAELTAAQVRVAAAAYETAVGLTVPPEVIAENRAELMI	120
262	QY	LIATNLLGQNTPAIVANAEYGEWQAQAAAMFGYAAATATATATLPLEEAPEMTSAGG	321
		121 LIATNLLGQNTPAIVANAEYGEWQAQAAAMFGYAAATATATATLPLEEAPEMTSAGG	180
322	QY	LLEQAAAAYEEASDTAAANQLMNNVPQALQALQPTGGTTPSSKLGGLWKTWTFGPHRSPISN	381
181	Db	LLEQAAAAYEEASDTAAANQLMNNVPQALQALQPTGGTTPSSKLGGLWKTWTFGPHRSPISN	240

QY 382 MVSMAHNSMTNSGVSMTNTLSMLKGFAPAPAAQAQVOTAAQNGVRAMSSLSGSSG 441
Db 241 MVSMAHNSMTNSGVSMTNTLSMLKGFAPAPAAQAQVOTAAQNGVRAMSSLSGSSG 300
QY 442 LGGVAAALGRAASVGSLSVFOAANAANOAVTPAARALPLTSLTSAARPGQMLGLPV 501
Db 301 LGGVAAALGRAASVGSLSVFOAANAANOAVTPAARALPLTSLTSAARPGQMLGLPV 360
QY 502 QMGARAGGSLGVLVPRPPRYVPHSPAAG 532
Db 361 QMGARAGGSLGVLVPRPPRYVPHSPAAG 391

RESULT 9
US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond S.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-107

Query Match 52.7%; Score 1944; DB 4; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134; 1; Indels 0; Gaps 0;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSWIG 201
Db 1 MVDGALPPEINSARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSWIG 60
QY 202 SSAGLVAAASPVVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
Db 61 SSAGLVAAASPVVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

QY 262 LIATNLGONTFAIIVNEAEYGEWMAQDAAMFGYAAATATATATLTLFFEPAPMTSAGG 321
Db 121 LIATNLGONTFAIIVNEAEYGEWMAQDAAMFGYAAATATATATLTLFFEPAPMTSAGG 180
QY 322 LLEQAAAVVEASDTAAANQLMNNVFOALQOLAQPTQGTTPSSKLGGLWKTVPKRSPTSN 381
Db 181 LLEQAAAVVEASDTAAANQLMNNVFOALQOLAQPTQGTTPSSKLGGLWKTVPKRSPTSN 240
QY 382 MVSMAHNSMTNSGVSMTNTLSMLKGFAPAPAAQAQVOTAAQNGVRAMSSLSGSSG 441
Db 241 MVSMAHNSMTNSGVSMTNTLSMLKGFAPAPAAQAQVOTAAQNGVRAMSSLSGSSG 300
QY 442 LGGVAAALGRAASVGSLSVFOAANAANOAVTPAARALPLTSLTSAARPGQMLGLPV 501
Db 301 LGGVAAALGRAASVGSLSVFOAANAANOAVTPAARALPLTSLTSAARPGQMLGLPV 360
QY 502 QMGARAGGSLGVLVPRPPRYVPHSPAAG 532
Db 361 QMGARAGGSLGVLVPRPPRYVPHSPAAG 391

RESULT 10
US-08-818-112-111
; Sequence 111, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond S.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-112-111

Query Match 44.8%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSWIG 201
Db 142 MVDGALPPEINSARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSWIG 201

Db 1 VVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFOQVVMGLTTGSGWIG 60
QY 202 SSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
Db 61 SSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPPEEAPMTSAGG 321
Db 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPPEEAPMTSAGG 180
QY 322 LLEQAAAVEASDTAAANQLMNNVPOALQOQAOTQGTTPSSKLGGLWKTVSPHRSPLSN 381
Db 181 LLEQAAVEAIDTAAANQLMNNVPOALQOQAOTQGTTPSSKLGGLWKTVSPHRSPLSN 240
QY 382 MVSMAHHMTNSGVSMTNTLSMLKGFAPAPAAQVOTAAQNGVRAMS-----LGSSL 437
Db 241 IVSMLNNHVSMTNSGVSMASTLHSMKGFAPAAA-QAVETAAQNGVQAMSSLGSLGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVQCAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGAGVAAANLGRAASVGSLSVQCAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 359
QY 498 GLPVGMQGARAG--GGLSGVLVPPRPVYVPHSPAAAG 532
Db 360 GLPLGLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396

RESULT 11

US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.41706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

Query Match

44.8%; Score 1652.5; DB 4; Length 396;

Best Local Similarity 84.9%; Pred. No. 6.2e-113; Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 142 MVDFFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFOQVVMGLTTGSGWIG 201
Db 1 VVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFOQVVMGLTTGSGWIG 60
QY 202 SSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
Db 61 SSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEEAPMTSAGG 321
Db 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEEAPMTSAGG 180
QY 322 LLEQAAAVEASDTAAANQLMNNVPOALQOQAOTQGTTPSSKLGGLWKTVSPHRSPLSN 381
Db 181 LLEQAAVEAIDTAAANQLMNNVPOALQOQAOTQGTTPSSKLGGLWKTVSPHRSPLSN 240
QY 382 MVSMAHHMTNSGVSMTNTLSMLKGFAPAPAAQVOTAAQNGVRAMS-----LGSSL 437
Db 241 IVSMLNNHVSMTNSGVSMASTLHSMKGFAPAAA-QAVETAAQNGVQAMSSLGSLGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVQCAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGAGVAAANLGRAASVGSLSVQCAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 359
QY 498 GLPVGMQGARAG--GGLSGVLVPPRPVYVPHSPAAAG 532
Db 360 GLPLGLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396

RESULT 12

US-09-056-556-111
; Sequence 111, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-056-556-111

Query Match	44.8%;	Score	1652.5;	DB	4;	Length	396;
Best Local Similarity	84.9%;	Pred.	No. 6.2e-113;				
Matches	337;	Conservative	19;	Mismatches	34;	Indels	7; Gaps
QY	142	MVDFCALPPEINSRMYAGPGSASLVAQAQMDSVASDLFSAASAFOQVVWGLTVGSWG	201				
Db	:	: : : : : :	:				
QY	202	SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPFPVIAENRASLMI	261				
Db	:	: : : : : :	:				
QY	61	SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPFPVIAENRASLMI	120				
Db	:	: : : : : :	:				
QY	262	LIAIATLLGONTPAIVANEAEGYGEWAQDAAMFGVAAATATATATLLPFEPAPMTCSGG	321				
Db	:	: : : : : :	:				
QY	121	LIAIATLLGONTPAIVANEAEGYGEWAQDAAMFGVAAATAATALLFFEDAPLITNPGG	180				
Db	:	: : : : : :	:				
QY	322	LLEQAAAVEEASDTAAANOLMNNVPOALQOLAQTQGTTPTPSKLGGLWKTVSPHRSPISN	381				
Db	:	: : : : : :	:				
QY	181	LLEQAVAVEAIDTAANAQLMNNVPOALQOLAQTPKSIWPFDQLSELWKALSPLHSPLSN	240				
Db	:	: : : : : :	:				
QY	382	MYSMANHHKMNTSGVSMVTNTLSMKUKGFAPAAAAQAVOTAAONGVRAMS - - - LGSLL	437				
Db	:	: : : : : :	:				
QY	241	IVSMIANHVSMNTSGVSMASTLHSLMKGFAPAAA-QAVETAONGVOAMSLGSQLGSSL	299				
Db	:	: : : : : :	:				
QY	438	GSSGLGGGAANTIGRAASVGSLSVPOAWAANCAVTPAARALPLTSLTSAERGPQGMILG	497				
Db	:	: : : : : :	:				
QY	300	GSSGLGGGAANTIGRAASVGSLSVPOAWAANCAVTPAARALPLTSLTSAEQATPHMG	359				
Db	:	: : : : : :	:				
QY	498	GLPVQMGABAG--GGLSGVLRVPPRPYPMPHSPAAG	532				
Db	:	: : : : : :	:				
QY	360	GLPVGOLITNSGCCGFGGVSNALRWPRAYVYWPVPVPAAG	396				
Db	:	: : : : : :	:				

RESULT 13
 US-09-072-596-106
 : Sequence 106, Application US/09072596
 : Patent No. 6458366
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Reed, Steven G.
 : APPLICANT: Skeiky, Yasir A.W.
 : APPLICANT: Dillon, Davin C.
 : APPLICANT: Campos-Neto, Antonia
 : APPLICANT: Houghton, Raymond
 : APPLICANT: Vedvick, Thomas S.
 : APPLICANT: Twardzik, Daniel R.
 : APPLICANT: Lodes, Michael J.
 : APPLICANT: Hendrickson, Ronald C.
 :
 : TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 : TUBERCULOSIS
 :
 : NUMBER OF SEQUENCES: 350
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: SEED AND BERRY LLP
 : STREET: 6300 Columbia Center, 701 Fifth Avenue
 : CITY: Seattle
 : STATE: Washington
 : COUNTRY: USA
 :
 : ZIP: 98104-7092
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/072,596
 : FILING DATE: 05-MAY-1998
 : CLASSIFICATION:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Maki, David J.
 : REGISTRATION NUMBER: 31,392
 : REFERENCE/DOCKET NUMBER: 210121.417C9
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (206) 622-4900
 : TELEFAX: (206) 682-6031
 : INFORMATION FOR SEQ ID NO: 106:

RESULT 14
US-09-072-967-111
; Sequence 111, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-967-111

Query Match 44.8%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

Qy 142 MVDGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSGWIG 201
Db 1 VVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Qy 202 SSAGLMVAASPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 261
Db 61 SSAGLMVAASPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
Qy 262 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPENTSAGG 321
Db 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPENTSAGG 180
Qy 322 LLEQAAVEEASDTAAANQMLNNVPAALQQLAQTGTPSKLGGIWKTVSPHRSPISN 381
Db 181 LLEQAAVEEASDTAAANQMLNNVPAALQQLAQTGTPSKLGGIWKTVSPHRSPISN 240
Qy 382 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAPAAARQAVTAAQNGVRAMSS----LGSSL 437
Db 241 IVSMNLNHSMTNSGVSMNTLSSMLKGFAPAPAAARQAVTAAQNGVRAMSS----LGSSL 299
Qy 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABERGPGQMLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABERGPGQMLG 359
Qy 498 GLPVQMGARAG--CGLSGLVLRPPRYMHPSPAAAG 532
Db 360 GLPLGOLTNSGGFGGVGNALRMPPRAYMHPVPAAG 396

RESULT 15
US-08-818-112-109
Sequence 109, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-818-112-109

Query Match 40.3%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.3e-101;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

Qy 142 MVDGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSGWIG 201
Db 1 VVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Qy 202 SSAGLMVAASPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 261
Db 61 SSAGLMVAASPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
Qy 262 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPENTSAGG 321
Db 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPENTSAGG 180
Qy 322 LLEQAAVEEASDTAAANQMLNNVPAALQQLAQTGTPSKLGGIWKTVSPHRSPISN 381
Db 181 LLEQAAVEEASDTAAANQMLNNVPAALQQLAQTGTPSKLGGIWKTVSPHRSPISN 240
Qy 382 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAPAAARQAVTAAQNGVRAMSS----LGSSL 437
Db 241 VSSIANNHSMNTNSGVSMNTLSSMLKGFAPAPAAARQAVTAAQNGVRAMSS----LGSSL 299
Qy 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABERGPGQMLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABERGPGQMLG 359

Search completed: June 30, 2004, 16:55:22
Job time : 24.1761 secs

The first part of the report deals with the general situation of the country and the progress of the work. It is followed by a detailed account of the various projects and the results obtained. The report concludes with a summary of the work done and the conclusions reached.

The second part of the report deals with the financial aspects of the work. It gives a detailed account of the income and expenditure of the organization and shows how the work has been financed.

The third part of the report deals with the personnel of the organization. It gives a list of the staff and describes their duties and the work they have done.

The fourth part of the report deals with the future of the organization. It gives a plan of the work to be done in the next year and describes the measures to be taken to improve the organization.

The fifth part of the report deals with the general conclusions of the work. It gives a summary of the main findings of the report and describes the measures to be taken to improve the organization.

The sixth part of the report deals with the general conclusions of the work. It gives a summary of the main findings of the report and describes the measures to be taken to improve the organization.

The seventh part of the report deals with the general conclusions of the work. It gives a summary of the main findings of the report and describes the measures to be taken to improve the organization.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:52:58 ; Search time 56.8458 Seconds
(without alignments)
3625.462 Million cell updates/sec

Title: US-09-597-796C-12

Perfect score: 3686

Sequence: 1 MHHHHTAASDNFQLSQGG.....SGGPVAVGLGQVGMNTAAS 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3686	100.0	729	9	US-09-287-849-2
2	3686	100.0	729	12	US-09-886-349A-16
3	3686	100.0	729	14	US-10-359-460-2
4	3686	100.0	729	14	US-10-098-732A-16
5	3686	100.0	729	15	US-10-359-459-2
6	3680	99.8	729	15	US-10-369-983-21
7	3680	99.8	813	15	US-10-369-983-15
8	3680	99.8	825	15	US-10-369-983-14
9	3680	99.8	875	15	US-10-369-983-13
10	3680	99.8	930	14	US-10-098-732A-65
11	3680	99.8	930	15	US-10-369-983-12
12	3680	99.8	1016	15	US-10-369-983-18
13	3680	99.8	1022	15	US-10-369-983-17
14	3680	99.8	1154	15	US-10-369-983-16
15	3677	99.8	729	12	US-09-886-349A-18

16	3677	99.8	729	14	US-10-098-732A-18	Sequence 18, Appl
17	3677	99.8	729	15	US-10-369-983-22	Sequence 22, Appl
18	2941	79.8	596	9	US-09-287-849-26	Sequence 26, Appl
19	2941	79.8	596	12	US-09-886-349A-20	Sequence 20, Appl
20	2941	79.8	596	14	US-10-359-460-25	Sequence 26, Appl
21	2941	79.8	596	14	US-10-098-732A-20	Sequence 20, Appl
22	2637	71.5	1010	15	US-10-369-983-4	Sequence 2, Appl
23	2631	71.4	723	15	US-10-369-983-2	Sequence 22, Appl
24	1967.5	53.4	600	9	US-09-287-849-22	Sequence 22, Appl
25	1967.5	53.4	600	14	US-10-359-460-22	Sequence 14, Appl
26	1944	52.7	391	12	US-09-886-349A-14	Sequence 22, Appl
27	1944	52.7	391	14	US-10-193-002-102	Sequence 102, App
28	1944	52.7	391	14	US-10-084-843-107	Sequence 107, App
29	1944	52.7	391	14	US-10-098-732A-14	Sequence 14, Appl
30	1929	52.3	391	12	US-09-872-186-8	Sequence 8, Appl
31	1652.5	44.8	396	14	US-10-193-002-106	Sequence 106, App
32	1652.5	44.8	396	14	US-10-084-843-111	Sequence 111, App
33	1583	42.9	393	12	US-10-282-122A-62455	Sequence 62455, A
34	1583	42.9	393	12	US-10-282-122A-64892	Sequence 64892, A
35	1486.5	40.3	359	14	US-10-193-002-104	Sequence 104, App
36	1486.5	40.3	359	14	US-10-084-843-109	Sequence 109, App
37	1184	32.1	358	9	US-09-287-849-8	Sequence 8, Appl
38	1184	32.1	358	14	US-10-359-460-8	Sequence 8, Appl
39	1182	32.1	263	12	US-09-886-349A-12	Sequence 12, Appl
40	1182	32.1	263	14	US-10-193-002-92	Sequence 92, Appl
41	1182	32.1	263	14	US-10-084-843-91	Sequence 91, Appl
42	1182	32.1	263	14	US-10-098-732A-12	Sequence 12, Appl
43	989	26.8	355	9	US-09-712-363-161	Sequence 161, App
44	987	26.8	330	12	US-09-886-349A-4	Sequence 4, Appl
45	987	26.8	330	14	US-10-098-732A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Protiens of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-2

Query Match 100.0%; Score 3686; DB 9; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.5e-244;

Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEHHHHTAASDNFQSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MEHHHHTAASDNFQSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGGARVQVRVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
DB 61 NNGGARVQVRVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120

QY 121 TKSGETTGNVTIAGPPAPFMDVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSGETTGNVTIAGPPAPFMDVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTACQAEITAAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTACQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPIAENRAELMILLIATNLLGONTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPIAENRAELMILLIATNLLGONTAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPPEEAPMTSAGLLEQAAAVVEASDTAAANQLMNVPOALQQLAQTQGT 360
DB 301 ATATATLLPPEEAPMTSAGLLEQAAAVVEASDTAAANQLMNVPOALQQLAQTQGT 360

QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420

QY 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAANCAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAANCAVTPAARALP 480

QY 481 LTSLSAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNAVGAGTGVIDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNAVGAGTGVIDPNGVLTNNHVA 600

QY 601 GATDINAFSVSGQTYGVVDVVGVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVVDVVGVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660

QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720

QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729

RESULT 2

US-09-886-349A-16
; Sequence 16, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; APPLICANT: Alderson, Mark
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 16
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
US-09-886-349A-16

Query Match 100.0%; Score 3686; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.5e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEHHHHTAASDNFQSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MEHHHHTAASDNFQSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGGARVQVRVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
DB 61 NNGGARVQVRVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120

QY 121 TKSGETTGNVTIAGPPAPFMDVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSGETTGNVTIAGPPAPFMDVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTACQAEITAAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTACQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPIAENRAELMILLIATNLLGONTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPIAENRAELMILLIATNLLGONTAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPPEEAPMTSAGLLEQAAAVVEASDTAAANQLMNVPOALQQLAQTQGT 360
DB 301 ATATATLLPPEEAPMTSAGLLEQAAAVVEASDTAAANQLMNVPOALQQLAQTQGT 360

QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420

QY 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAANCAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAANCAVTPAARALP 480

QY 481 LTSLSAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNAVGAGTGVIDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNAVGAGTGVIDPNGVLTNNHVA 600

QY 601 GATDINAFSVSGQTYGVVDVVGVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVVDVVGVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660

QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720

QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729

RESULT 3

US-10-359-460-2
; Sequence 2, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match
Best Local Similarity 100.0%; Score 3686; DB 14; Length 729;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Qy 61 NNGGARVQVRVGSAPASLSGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGGARVQVRVGSAPASLSGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Qy 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSGWIGSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWIGSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
Qy 241 TAYGLTVPPVPIAENRAELMILITNLGQNTPAIVNEAEYGENWAOAAAFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMILITNLGQNTPAIVNEAEYGENWAOAAAFGYAAAT 300
Qy 301 ATATATLLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGT 360
Qy 361 PSSKLGGLWKTVPSPHRSPTSNMVSNNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPTSNMVSNNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
Qy 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
Qy 481 LTSLSAERGFQGMGLGVQCMGARAGGGLSGVLVPPRPYVPHSPAGDIIAPPALS 540
Db 481 LTSLSAERGFQGMGLGVQCMGARAGGGLSGVLVPPRPYVPHSPAGDIIAPPALS 540

Qy 541 ODRFADFPALDPSAMVAQVGPVNTKLGYNNAVGAGTGIVDPNGVVLNNHVLIA 600
Db 541 ODRFADFPALDPSAMVAQVGPVNTKLGYNNAVGAGTGIVDPNGVVLNNHVLIA 600
Qy 601 GATDINAFSVSGQTYGVDVVDYDRTDQDVAVLQRGAGGLPSAAIGGVAVGPPVAMGN 660
Db 601 GATDINAFSVSGQTYGVDVVDYDRTDQDVAVLQRGAGGLPSAAIGGVAVGPPVAMGN 660
Qy 661 SGGGGTFRVAVGRVAVAGTQVQASDLSLTGABETLNGLIQFDAAIOPGDSGGPVVNLGQ 720
Db 661 SGGGGTFRVAVGRVAVAGTQVQASDLSLTGABETLNGLIQFDAAIOPGDSGGPVVNLGQ 720
Qy 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 4
US-10-098-732A-16
Sequence 16, Application US/10098732A
Publication No. US20030175294A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guiderian, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
TITLE OF INVENTION: Leishmania Antigen
FILE REFERENCE: 014058-012010US
CURRENT APPLICATION NUMBER: US/10/098,732A
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: US 60/275,837
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 16
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
OTHER INFORMATION: Protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
OTHER INFORMATION: fusion)
US-10-098-732A-16

Query Match
Best Local Similarity 100.0%; Score 3686; DB 14; Length 729;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Qy 61 NNGGARVQVRVGSAPASLSGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGGARVQVRVGSAPASLSGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Qy 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSGWIGSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWIGSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
Qy 241 TAYGLTVPPVPIAENRAELMILITNLGQNTPAIVNEAEYGENWAOAAAFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMILITNLGQNTPAIVNEAEYGENWAOAAAFGYAAAT 300
Qy 301 ATATATLLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGT 360

361 PSSKLGGLWKTVPSPHRSPISNMWSMANNHSMNTNSGVSMTNTLSSMLKGFAPAPAAARQAVQ 420
Db |
361 PSSKLGGLWKTVPSPHRSPISNMWSMANNHSMNTNSGVSMTNTLSSMLKGFAPAPAAARQAVQ 420
Qy |
421 TAAQNGVRAMSSSLGSSSLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
Db |
421 TAAQNGVRAMSSSLGSSSLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
Qy |
481 LTSLSAERPGOMLGGGLPVGMGARAGGGLSGVLRVPPRYPMHSPHSPAAGDIAPPALS 540
Db |
481 LTSLSAERPGOMLGGGLPVGMGARAGGGLSGVLRVPPRYPMHSPHSPAAGDIAPPALS 540
Qy |
541 QDRFADFPALPLDPSAMVAQVGVVNTNKLGNNAVAGAGTGVLDPNGLVLTNNHVA 600
Db |
541 QDRFADFPALPLDPSAMVAQVGVVNTNKLGNNAVAGAGTGVLDPNGLVLTNNHVA 600
Qy |
601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db |
601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Qy |
661 SGGGGTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGSGGPPVNNGLGQ 720
Db |
661 SGGGGTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGSGGPPVNNGLGQ 720
Qy |
721 VVGMMNTAAS 729
Db |
721 VVGMMNTAAS 729
Qy |
RESULT 5
US-10-359-459-2
; Sequence 2, Application US/10359459
; Publication No. US20040013677A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/10/359,459
; CURRENT FILING DATE: 2003-02-05
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-459-2
Query Match 100.0%; Score 3686; DB 15; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.5e-244; Indels 0; Gaps 0;
Matches 729; Conservative 0; Mismatches 0;
Qy 1 MHHHHHTAASDNFQSLQSGGQGFAPIGQAMATAGQIRSGGSPVTHIGTAFGLGLGVD 60
Db 1 MHHHHHTAASDNFQSLQSGGQGFAPIGQAMATAGQIRSGGSPVTHIGTAFGLGLGVD 60
Qy 61 NNGNGARVQRVVGSGAPASISIGTGDVITAVDGAPINSATAMADALNGHHPGDVSVTWQ 120
Db 61 NNGNGARVQRVVGSGAPASISIGTGDVITAVDGAPINSATAMADALNGHHPGDVSVTWQ 120
Qy 121 TKSGGTGTGNVTLAEQPPAEFMDVFGALPPEINSAARMYAGPGSASLVAAQWDSVASDL 180
Db 121 TKSGGTGTGNVTLAEQPPAEFMDVFGALPPEINSAARMYAGPGSASLVAAQWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSSWIGSSAGLVAASPYVAMSVTAGQAEILTAQVRVAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSSWIGSSAGLVAASPYVAMSVTAGQAEILTAQVRVAAAYE 240

181 FSAASAFQSVVWGLTVGSSWIGSSAGLVAASPYVAMSVTAGQAEILTAQVRVAAAYE 240
Qy |
241 TAYGLTVPPPVIAENRAELMILTIATNLLGONTPAIAVNEAEYEGEMWAQDAAMFGYAAAT 300
Db |
241 TAYGLTVPPPVIAENRAELMILTIATNLLGONTPAIAVNEAEYEGEMWAQDAAMFGYAAAT 300
Qy |
301 ATATATLLPPEAEPMTSAGGLLEQAAAEEASDTAAANOLMNNVPOALQOLAOPTQGT 360
Db |
301 ATATATLLPPEAEPMTSAGGLLEQAAAEEASDTAAANOLMNNVPOALQOLAOPTQGT 360
Qy |
361 PSSKLGGLWKTVPSPHRSPISNMWSMANNHSMNTNSGVSMTNTLSSMLKGFAPAPAAARQAVQ 420
Db |
361 PSSKLGGLWKTVPSPHRSPISNMWSMANNHSMNTNSGVSMTNTLSSMLKGFAPAPAAARQAVQ 420
Qy |
421 TAAQNGVRAMSSSLGSSSLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
Db |
421 TAAQNGVRAMSSSLGSSSLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
Qy |
481 LTSLSAERPGOMLGGGLPVGMGARAGGGLSGVLRVPPRYPMHSPHSPAAGDIAPPALS 540
Db |
481 LTSLSAERPGOMLGGGLPVGMGARAGGGLSGVLRVPPRYPMHSPHSPAAGDIAPPALS 540
Qy |
541 QDRFADFPALPLDPSAMVAQVGVVNTNKLGNNAVAGAGTGVLDPNGLVLTNNHVA 600
Db |
541 QDRFADFPALPLDPSAMVAQVGVVNTNKLGNNAVAGAGTGVLDPNGLVLTNNHVA 600
Qy |
601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db |
601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Qy |
661 SGGGGTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGSGGPPVNNGLGQ 720
Db |
661 SGGGGTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGSGGPPVNNGLGQ 720
Qy |
721 VVGMMNTAAS 729
Db |
721 VVGMMNTAAS 729
Qy |
RESULT 6
US-10-369-983-21
; Sequence 21, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 21
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72F
US-10-369-983-21
Query Match 99.8%; Score 3680; DB 15; Length 729;
Best Local Similarity 99.9%; Pred. No. 1.9e-243; Indels 0; Gaps 0;
Matches 728; Conservative 0; Mismatches 1;
Qy 1 MHHHHHTAASDNFQSLQSGGQGFAPIGQAMATAGQIRSGGSPVTHIGTAFGLGLGVD 60
Db 1 MHHHHHTAASDNFQSLQSGGQGFAPIGQAMATAGQIRSGGSPVTHIGTAFGLGLGVD 60
Qy 61 NNGNGARVQRVVGSGAPASISIGTGDVITAVDGAPINSATAMADALNGHHPGDVSVTWQ 120

Db 61 NNGNGARVQVVGSAAPASLGIISTGDVITAVDGAIPNSATAMADALNGHHGPDVLSVTWQ 120
QY 121 TYSGGTRTGNVTLAGCPPAEFWDFGALPPEINSAARMVAGPQASLSLVAAQWMDSVASDL 180
Db 121 TYSGGTRTGNVTLAGCPPAEFWDFGALPPEINSAARMVAGPQASLSLVAAQWMDSVASDL 180
QY 181 FSAASAFQSVWGLTVGSGWISGAGLMVAASAPYVAMSVTAGOAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVWGLTVGSGWISGAGLMVAASAPYVAMSVTAGOAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHSPISSNNMVMANNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHSPISSNNMVMANNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGAGLVGQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPALS 540
Db 421 TAAQNGVRAMSSLGSSLGSSGAGLVGQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPALS 600
Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPALS 600
QY 541 QDRFADFPALPLDPSAMVAQVGPVNNINTKLGYNNAVGAGTGIVIDPNGVVLNNHVA 660
Db 541 QDRFADFPALPLDPSAMVAQVGPVNNINTKLGYNNAVGAGTGIVIDPNGVVLNNHVA 660
QY 601 GATDINAFSVGSGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 720
Db 601 GATDINAFSVGSGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 7

US-10-369-983-15
; Sequence 15, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB81F (MTB72F-DFV)
US-10-369-983-15

Query Match 99.8%; Score 3680; DB 15; Length 813;
Best Local Similarity 99.9%; Pred. No. 2.2e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHHTAASDNFQLSGGGQGFAPITGQAMAIAGQIRSGGSPPTVHIGTAFILGLGVD 60
Db 1 MHHHHHHTAASDNFQLSGGGQGFAPITGQAMAIAGQIRSGGSPPTVHIGTAFILGLGVD 60
QY 61 NNGNGARVQVVGSAAPASLGIISTGDVITAVDGAIPNSATAMADALNGHHGPDVLSVTWQ 120
Db 61 NNGNGARVQVVGSAAPASLGIISTGDVITAVDGAIPNSATAMADALNGHHGPDVLSVTWQ 120
QY 121 TKSOGTGTGNVTLAGPPAEFWDGALPPEINSAARMVAGPQASLSLVAAQWMDSVASDL 180
Db 121 TKSOGTGTGNVTLAGPPAEFWDGALPPEINSAARMVAGPQASLSLVAAQWMDSVASDL 180
QY 181 FSAASAFQSVWGLTVGSGWISGAGLMVAASAPYVAMSVTAGOAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVWGLTVGSGWISGAGLMVAASAPYVAMSVTAGOAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHSPISSNNMVMANNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHSPISSNNMVMANNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGAGLVGQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPALS 480
Db 421 TAAQNGVRAMSSLGSSLGSSGAGLVGQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPALS 480
QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPVNNINTKLGYNNAVGAGTGIVIDPNGVVLNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPVNNINTKLGYNNAVGAGTGIVIDPNGVVLNNHVA 600
QY 601 GATDINAFSVGSGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVGSGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 8

US-10-369-983-14
; Sequence 14, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15

```
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 825
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB83F (MTB72F-MTI)
US-10-369-983-14

Query Match      99.8%; Score 3680; DB 15; Length 825;
Best Local Similarity 99.8%; Pred. No. 2.3e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGLGVVD 60
QY 61 NNGNGARVQVWGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
DB 61 NNGNGARVQVWGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
QY 121 TKSGGTGTGNTVLAEGPPAEFMDFGALPPEINSARYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGGTGTGNTVLAEGPPAEFMDFGALPPEINSARYAGPGSASLVAAQMWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVPIAENRAELMILITATNLGONTPAIAVNEAEYGEWMAQAAAFVGAAT 300
DB 241 TAYGLTVPPVPIAENRAELMILITATNLGONTPAIAVNEAEYGEWMAQAAAFVGAAT 300
QY 301 ATATATLLPPEEAPMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQAQPTQGT 360
DB 301 ATATATLLPPEEAPMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMSWNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMSWNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLRVPPRYPMPSHSPAAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLRVPPRYPMPSHSPAAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVQVNNITKLYNNNAVAGTGVIDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGVQVNNITKLYNNNAVAGTGVIDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGGQTYGVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGGQTYGVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVPGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
DB 661 SGGGGTTPRAVPGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729
```

RESULT 9
US-10-369-983-13
; Sequence 13, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:

```
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIORITY FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 875
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-Erd14)
US-10-369-983-13
```

```
Query Match      99.8%; Score 3680; DB 15; Length 875;
Best Local Similarity 99.8%; Pred. No. 2.4e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGLGVVD 60
QY 61 NNGNGARVQVWGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
DB 61 NNGNGARVQVWGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
QY 121 TKSGGTGTGNTVLAEGPPAEFMDFGALPPEINSARYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGGTGTGNTVLAEGPPAEFMDFGALPPEINSARYAGPGSASLVAAQMWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVPIAENRAELMILITATNLGONTPAIAVNEAEYGEWMAQAAAFVGAAT 300
DB 241 TAYGLTVPPVPIAENRAELMILITATNLGONTPAIAVNEAEYGEWMAQAAAFVGAAT 300
QY 301 ATATATLLPPEEAPMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQAQPTQGT 360
DB 301 ATATATLLPPEEAPMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMSWNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMSWNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLRVPPRYPMPSHSPAAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLRVPPRYPMPSHSPAAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVQVNNITKLYNNNAVAGTGVIDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGVQVNNITKLYNNNAVAGTGVIDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGGQTYGVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGGQTYGVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVPGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
DB 661 SGGGGTTPRAVPGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
```



```
QY 721 VVGMNTAAS 729
Db 721 VVGMNTAAS 729

RESULT 10
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication NO. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guiderian, Jeff
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F-MAPS
; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Ra12-TbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-65

Query Match 99.8%; Score 3680; DB 14; Length 930;
Best Local Similarity 99.9%; Pred. No. 2.6e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 61 NNGNGARVQVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASIVAAQMWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASIVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSMIGSSAGLMVAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSMIGSSAGLMVAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240

QY 241 TAYGLTVPPVIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEAPENTTSAGGLEQAAAEEASDTAAANQLMNNVPAQLQLAQPTQGT 360
Db 301 ATATATLLPFEAPENTTSAGGLEQAAAEEASDTAAANQLMNNVPAQLQLAQPTQGT 360

QY 361 PSSKLGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420

QY 421 TAAQNGVRAMSSGSSIGSSGLGGGVAANI GRAASVGSLSVPCAWAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSGSSIGSSGLGGGVAANI GRAASVGSLSVPCAWAAANQAVTPAARALP 480

QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGSLGVLVRPPRPPVMPHSPAAAGDIAPPALS 540
Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGSLGVLVRPPRPPVMPHSPAAAGDIAPPALS 540
```

```
QY 541 QDRFADPPALPLDPSAMVAQVGPVNVNITKLYNNNAVAGGTGIVIDPKNVLTNNHVL 600
Db 541 QDRFADPPALPLDPSAMVAQVGPVNVNITKLYNNNAVAGGTGIVIDPKNVLTNNHVL 600

QY 601 GATDINAFSVGSGQTYGVDDVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGPPVAMGN 660
Db 601 GATDINAFSVGSGQTYGVDDVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGPPVAMGN 660

QY 661 SGGGGTTPRAVPRVVALGQTVQASDSLTCABETLNGLIQFDAAIQBDSGGPVVNLGQ 720
Db 661 SGGGGTTPRAVPRVVALGQTVQASDSLTCABETLNGLIQFDAAIQBDSGGPVVNLGQ 720

QY 721 VVGMNTAAS 729
Db 721 VVGMNTAAS 729

RESULT 11
US-10-369-983-12
; Sequence 12, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guiderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: R95F (MTB72F-MAPS)
US-10-369-983-12

Query Match 99.8%; Score 3680; DB 15; Length 930;
Best Local Similarity 99.9%; Pred. No. 2.6e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 61 NNGNGARVQVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASIVAAQMWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASIVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSMIGSSAGLMVAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSMIGSSAGLMVAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240

QY 241 TAYGLTVPPVIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEAPENTTSAGGLEQAAAEEASDTAAANQLMNNVPAQLQLAQPTQGT 360
Db 301 ATATATLLPFEAPENTTSAGGLEQAAAEEASDTAAANQLMNNVPAQLQLAQPTQGT 360

QY 361 PSSKLGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
```

Db PSSKLGGLWKTVPSPHRSPIINMWSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAAAQAVQ 420
Qy TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALP 480
Db TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALP 480
Qy LTSLSAAERPGQMLGGLPVQNGARAGGGLSGVLVPRPYPYMPHSPAAAGDIAPPALS 540
Db LTSLSAAERPGQMLGGLPVQNGARAGGGLSGVLVPRPYPYMPHSPAAAGDIAPPALS 540
Qy QDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAVGAGTGIVDPNGVLTNNHVA 600
Db QDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAVGAGTGIVDPNGVLTNNHVA 600
Qy GATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db GATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Qy SGGGGTTPRAVPGRVAVLQVQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
Db SGGGGTTPRAVPGRVAVLQVQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
Qy VVGMMNTAAS 729
Db VVGMMNTAAS 729

RESULT 12

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 99.8%; Score 3680; DB 15; Length 1016;
Best Local Similarity 99.9%; Pred. No. 3e-243; Indels 0; Gaps 0;
Matches 728; Conservative 0; Mismatches 1;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGQIRSGGSPVHIGTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGQIRSGGSPVHIGTAFGLGVVD 60
Qy 61 NNGNGARVQRVVSAPAAASLGISTGDVITAVDGAIPNSATAMADALNGHHPGDVISVTWQ 120
Db 61 NNGNGARVQRVVSAPAAASLGISTGDVITAVDGAIPNSATAMADALNGHHPGDVISVTWQ 120
Qy 121 TKSGGTRTGNVTLAEPPFAEFVDFGALPPEINSAEMVAGPSSASIVAAQWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEPPFAEFVDFGALPPEINSAEMVAGPSSASIVAAQWDSVASDL 180
Qy 181 FSAASAPQSVVWGLTVGSSWITGSSAGLMVAASPYVAMSVTACQAEELTAQVRVAAAAVE 240
Db 181 FSAASAPQSVVWGLTVGSSWITGSSAGLMVAASPYVAMSVTACQAEELTAQVRVAAAAVE 240

Qy 241 TAYGLTVPVPPVIAENRAELMILITATNLGQNTPAIVNNEAYGEMWAQDAAMFGYAAAT 300
Db 241 TAYGLTVPVPPVIAENRAELMILITATNLGQNTPAIVNNEAYGEMWAQDAAMFGYAAAT 300
Qy 301 ATATATLPPFEAPMTSAGGLLEQAAAVBEASDTAAANQLMNNVPOALQOLAQTOGTT 360
Db 301 ATATATLPPFEAPMTSAGGLLEQAAAVBEASDTAAANQLMNNVPOALQOLAQTOGTT 360
Qy PSSKLGGLWKTVPSPHRSPIINMWSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAAAQAVQ 420
Db PSSKLGGLWKTVPSPHRSPIINMWSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAAAQAVQ 420
Qy TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALP 480
Db TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALP 480
Qy LTSLSAAERPGQMLGGLPVQNGARAGGGLSGVLVPRPYPYMPHSPAAAGDIAPPALS 540
Db LTSLSAAERPGQMLGGLPVQNGARAGGGLSGVLVPRPYPYMPHSPAAAGDIAPPALS 540
Qy QDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAVGAGTGIVDPNGVLTNNHVA 600
Db QDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAVGAGTGIVDPNGVLTNNHVA 600
Qy GATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db GATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Qy SGGGGTTPRAVPGRVAVLQVQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
Db SGGGGTTPRAVPGRVAVLQVQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
Qy VVGMMNTAAS 729
Db VVGMMNTAAS 729

RESULT 13

US-10-369-983-17
; Sequence 17, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1022
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB102FM2F (MTB102FTW2, MTB72F-HTCC#1)
US-10-369-983-17

Query Match 99.8%; Score 3680; DB 15; Length 1022;
Best Local Similarity 99.9%; Pred. No. 3e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGQIRSGGSPVHIGTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGQIRSGGSPVHIGTAFGLGVVD 60
Qy 61 NNGNGARVQRVVSAPAAASLGISTGDVITAVDGAIPNSATAMADALNGHHPGDVISVTWQ 120

Db 61 NNGGARVQRVVGSAAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
QY 121 TSKGGTRTGNVTLAEGPPAEFVDFGALPPEINSAEMVAGPSASIVAAQWMDVSADL 180
Db 121 TSKGGTRTGNVTLAEGPPAEFVDFGALPPEINSAEMVAGPSASIVAAQWMDVSADL 180
QY 181 FSAASAFQSVVWGLTVGSMIGSAGLWVAASPYVAMSVTGAQAEITAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSMIGSAGLWVAASPYVAMSVTGAQAEITAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVTAENRAELMILITATLLQNTPTAIVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVTAENRAELMILITATLLQNTPTAIVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMTNSGVMTNTLSSMLKGFAPAAAQVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMTNSGVMTNTLSSMLKGFAPAAAQVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQOMLGGLPVQMGARAGGSLGVLVPPRYVMPHSPAAGDIAPFALS 540
Db 481 LTSLSAAERGPQOMLGGLPVQMGARAGGSLGVLVPPRYVMPHSPAAGDIAPFALS 540
QY 541 QDRPADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGAGTGIVDPNGVVLTTNNHVA 600
Db 541 QDRPADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGAGTGIVDPNGVVLTTNNHVA 600
QY 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAALGGGVAVGEPVWAMGN 660
Db 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAALGGGVAVGEPVWAMGN 660
QY 661 SGGQGGTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDDAIOPGDSGGPVVWNLGQ 720
Db 661 SGGQGGTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDDAIOPGDSGGPVVWNLGQ 720
QY 721 VVGMMNTAAS 729
Db 721 VVGMMNTAAS 729

RESULT 14
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Gudarian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-ATCC#2)
US-10-369-983-16

Query Match 99.8%; Score 3680; DB 15; Length 1154;
Best Local Similarity 99.9%; Pred No. 3.5e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHEHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGTAFIAGLGVVD 60
Db 1 MHHEHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGTAFIAGLGVVD 60
QY 61 NNGGARVQRVVGSAAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGGARVQRVVGSAAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
QY 121 TSKGGTRTGNVTLAEGPPAEFVDFGALPPEINSAEMVAGPSASIVAAQWMDVSADL 180
Db 121 TSKGGTRTGNVTLAEGPPAEFVDFGALPPEINSAEMVAGPSASIVAAQWMDVSADL 180
QY 181 FSAASAFQSVVWGLTVGSMIGSAGLWVAASPYVAMSVTGAQAEITAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSMIGSAGLWVAASPYVAMSVTGAQAEITAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVTAENRAELMILITATLLQNTPTAIVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVTAENRAELMILITATLLQNTPTAIVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMTNSGVMTNTLSSMLKGFAPAAAQVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMTNSGVMTNTLSSMLKGFAPAAAQVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQOMLGGLPVQMGARAGGSLGVLVPPRYVMPHSPAAGDIAPFALS 540
Db 481 LTSLSAAERGPQOMLGGLPVQMGARAGGSLGVLVPPRYVMPHSPAAGDIAPFALS 540
QY 541 QDRPADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGAGTGIVDPNGVVLTTNNHVA 600
Db 541 QDRPADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGAGTGIVDPNGVVLTTNNHVA 600
QY 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAALGGGVAVGEPVWAMGN 660
Db 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAALGGGVAVGEPVWAMGN 660
QY 661 SGGQGGTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDDAIOPGDSGGPVVWNLGQ 720
Db 661 SGGQGGTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDDAIOPGDSGGPVVWNLGQ 720
QY 721 VVGMMNTAAS 729
Db 721 VVGMMNTAAS 729

RESULT 15
US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20

Wed Jul 7 12:10:32 2004

; PRIOR APPLICATION NUMBER: US 60/265,737
 ; PRIOR FILING DATE: 2001-02-01
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: Patencin Ver. 2.1
 ; SEQ ID NO 18
 ; LENGTH: 729
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
 ; OTHER INFORMATION: (Ra12-TbHP-Ra35MutSA)
 US-09-886-349A-18

Query Match	99.8%	Score 3677	DB 12	Length 729
Best Local Similarity	99.7%	Pred. No. 3.1e-243		
Matches 727	Conservative 1	Mismatches 1	Indels 0	Gaps 0

Qy	1	MHHHHHTAASDNFQLSQGGGFAIPIGQMAIAGQIRSGGSPVHIGTAFGLGVVD	60
Db	1	MHHHHHTAASDNFQLSQGGGFAIPIGQMAIAGQIRSGGSPVHIGTAFGLGVVD	60
Qy	61	NNNGARVQVWGVSAPASIGTGTGVTAVDGPINSATATADALNGHHPGDVISVTWQ	120
Db	61	NNNGARVQVWGVSAPASIGTGTGVTAVDGPINSATATADALNGHHPGDVISVTWQ	120
Qy	121	TKSGGTRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL	180
Db	121	TKSGGTRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL	180
Qy	181	FSASAFQSVVWGLTVGWSWIGSSAGLMVAAASPVVAWMSVTAGCAELTAAQVRVAAAAYE	240
Db	181	FSASAFQSVVWGLTVGWSWIGSSAGLMVAAASPVVAWMSVTAGCAELTAAQVRVAAAAYE	240
Qy	241	TAYGLTVPPVPIAENRAELMILITNLGNTPAIAVNEAEYCEMWAQDAAMFGYAAAT	300
Db	241	TAYGLTVPPVPIAENRAELMILITNLGNTPAIAVNEAEYCEMWAQDAAMFGYAAAT	300
Qy	301	ATATATLLPPEEAPEMTSAGGLLEQAAVEASDTAAANQLMNNVPCALQOLAQPTQGT	360
Db	301	ATATATLLPPEEAPEMTSAGGLLEQAAVEASDTAAANQLMNNVPCALQOLAQPTQGT	360
Qy	361	PSSKLGGLWKTVPSPHRSPIENNVSMANNHMTNSGVSMNTILSSMLKGFAPAAARQAVQ	420
Db	361	PSSKLGGLWKTVPSPHRSPIENNVSMANNHMTNSGVSMNTILSSMLKGFAPAAARQAVQ	420
Qy	421	TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALP	480
Db	421	TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALP	480
Qy	481	LTSLSAERGPQGMGLGVQNGARAGGCLSGVLRVPPRVPVMPHSPAAGDIAPPALS	540
Db	481	LTSLSAERGPQGMGLGVQNGARAGGCLSGVLRVPPRVPVMPHSPAAGDIAPPALS	540
Qy	541	QDRFADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVGAGTGVIDPNGVLTNNHVA	600
Db	541	QDRFADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVGAGTGVIDPNGVLTNNHVA	600
Qy	601	GATDINAFSVSGQTYGVVDVDRCTODVAVLQIRGAGGLPSAAGGVAVGEPVVMGN	660
Db	601	GATDINAFSVSGQTYGVVDVDRCTODVAVLQIRGAGGLPSAAGGVAVGEPVVMGN	660
Qy	661	SGGQGTFRVAVGRTVQASDLSLTGAETLNGLIQFDDAIIQPGSGGFPVNVGLQ	720
Db	661	SGGQGTFRVAVGRTVQASDLSLTGAETLNGLIQFDDAIIQPGSGGFPVNVGLQ	720
Qy	721	VVGNTAAS 729	
Db	721	VVGNTAAS 729	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 19.0527 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796C-12
Perfect score: 3686
Sequence: 1 MHHEHHHTAASDNFQLSQGG.....SGGPVNVGLGVGMNTAAS 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1944	52.7	391	2 B70608	probable PPE prote
2	1656.5	44.9	396	2 H70741	probable PPE prote
3	1583	42.9	393	2 C70568	probable PPE prote
4	989	26.8	355	2 F70983	probable serine pr
5	778.5	21.1	361	2 S47170	hypothetical prote
6	775.5	21.0	393	2 G70929	probable PPE prote
7	754.5	20.5	354	2 A87242	probable secreted
8	753.5	20.4	409	2 A70932	probable PPE prote
9	737	20.0	423	2 B70931	probable PPE prote
10	734	19.9	421	2 H87056	PPE-family protein
11	733.5	19.9	403	2 H70931	probable PPE prote
12	702	19.0	408	2 G70925	probable PPE prote
13	700	19.0	413	2 F70560	probable PPE prote
14	697	18.9	391	2 F70625	probable PPE prote
15	689	18.7	468	2 B70932	probable PPE prote
16	687.5	18.7	463	2 C70931	probable PPE prote
17	674	18.3	380	2 A70646	probable PPE prote
18	670.5	18.2	394	2 G70881	probable PPE prote
19	664.5	18.0	385	2 H70503	probable PPE prote
20	635	17.2	350	2 F70929	probable PPE prote
21	633.5	17.2	365	2 F70929	probable PPE prote
22	615.5	16.7	402	2 A70882	probable PPE prote
23	605.5	16.4	423	2 C70582	probable PPE prote
24	601.5	16.3	391	2 D70922	probable PPE prote
25	597.5	15.2	394	2 A70504	probable PPE prote
26	584	15.8	406	2 E70675	probable PPE prote
27	570.5	15.5	391	2 A70663	probable PPE prote
28	520.5	14.1	3300	2 D70575	probable PPE prote
29	476.5	12.9	3716	2 E70969	probable PPE prote

30	473	12.8	180	2 G70834	probable PPE prote
31	463	12.6	655	2 A70931	probable PPE prote
32	462.5	12.5	580	2 G70570	probable PPE prote
33	458	12.4	3157	2 B70969	probable PPE prote
34	452	12.3	2523	2 F70846	probable PPE prote
35	451	12.2	963	2 B70524	probable PPE prote
36	450	12.2	678	2 A70762	probable PPE prote
37	449	12.2	346	2 H70874	probable PPE prote
38	447.5	12.1	582	2 F70675	probable PPE prote
39	443	12.0	479	2 D70676	probable PPE prote
40	443	12.0	552	2 D70604	probable PPE prote
41	442	12.0	487	2 E70830	probable PPE prote
42	437.5	11.9	615	2 E70663	probable PPE prote
43	429.5	11.7	645	2 F70825	probable PPE prote
44	426.5	11.6	1053	2 B70987	probable PPE prote
45	425.5	11.5	590	2 E70946	probable PPE prote

ALIGNMENTS

RESULT 1

B70608
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: B70608
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70608
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-391 <COL>
A;Cross-references: GB:Z93777; GB:AL123456; NID:G3261726; PIDN:CAB07839.1; PID:e311073; I
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: PPE

Query Match	52.7%	Score	1944	DB	2	Length	391
Best Local Similarity	99.7%	Pred. No.	4e-92				
Matches	390	Conservative	0	Mismatches	1	Indels	0
Gaps	0						
Qy	142	MVDFGALPPEINARMYAGPGSASLVAAACQMDVSASDLFSAASAFQSVVWGLTVGSWIG	201				
Db	1	MVDFGALPPEINARMYAGPGSASLVAAACQMDVSASDLFSAASAFQSVVWGLTVGSWIG	60				
Qy	202	SSAGLMVAASPYVAMSVTAGQALTAQVRVAAAAYETAYGLTVPPPIAENRAELMI	261				
Db	61	SSAGLMVAASPYVAMSVTAGQALTAQVRVAAAAYETAYGLTVPPPIAENRAELMI	120				
Qy	262	LIATNLGONTPTAINEAEYGEWMAQDAANFGYAAATATATATALLPPEEAPMTSAGG	321				
Db	121	LIATNLGONTPTAINEAEYGEWMAQDAANFGYAAATATATATALLPPEEAPMTSAGG	180				
Qy	322	LLEQAAVSEASDTAAANQLMNNVPQALQQAQPTQCTTPSSKLGGLWKTSPHRSPISN	381				
Db	181	LLEQAAVSEASDTAAANQLMNNVPQALQQAQPTQCTTPSSKLGGLWKTSPHRSPISN	240				
Qy	382	MVSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVOTAONGVRAMSSLGSSG	441				
Db	241	MVSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVOTAONGVRAMSSLGSSG	300				
Qy	442	LGCGVAAVSEASDTAAANQLMNNVPQALQQAQPTQCTTPSSKLGGLWKTSPHRSPISN	501				
Db	301	LGCGVAAVSEASDTAAANQLMNNVPQALQQAQPTQCTTPSSKLGGLWKTSPHRSPISN	360				
Qy	502	GQMGAPAGGLGVLVPPRPYPYMPHSPAAG	532				

Db 361 GQMGRAGGGLSGVLRVPRPPYVMPHSPAAG 391

RESULT 2

H70741

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: H70741

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sultston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: H70741

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-396 <COL>

A;Cross-references: GB:275555; GB:AL123456; NID:g2261608; PIDN:CAA99966.1; PID:e250360;

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: PPE

Query Match 44.9%; Score 1656.5; DB 2; Length 396;

Best Local Similarity 85.1%; Pred. No. 1.8e-77;

Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDGALPPPEINSGARMYVAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSMIG 201

Db 1 MVDGALPPPEINSGARMYVAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSMIG 60

QY 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261

Db 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120

QY 262 LIATNLGQNTPAIAVNEABYGMWAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 321

Db 121 LIATNLGQNTPAIAVNEABYGMWAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180

QY 322 LLEQAAVEEASDTAAANQLMNNVPOALQQAQPTGCTTPSSKLGGLWKTVPSPHSPISN 381

Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQQAQPTGCTTPSSKLGGLWKTVPSPHSPISN 240

QY 382 MVSNNHNSMTNSGVSMTNTLSMLKGFAPAAQAVOTAAQNGVRMSS---LGSSL 437

Db 438 GSSGLGGVAAANLGRAASVGLSVPOQAAWAAANQAVTPAARALPLTSLTSAERGPQMLG 497

Db 498 GLPVQMGARAG--GGLSGVLRVPRPPYVMPHSPAAG 532

Db 360 GLPLGQLTNSGGGGVSNALRMPPRAYVPRVPAAG 396

RESULT 3

C70568

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: C70568

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sultston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: C70568

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-355 <COL>

A;Cross-references: GB:296071; GB:AL123456; NID:g3242254; PIDN:CAB09453.1; PID:g2181967

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: pepA

C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypsin

Query Match 26.8%; Score 989; DB 2; Length 355;

Best Local Similarity 98.0%; Pred. No. 1.5e-43;

Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 530 AAGDIAPALSDQDFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIDPN 589

Db 28 APAQAAPALSDQDFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIDPN 87

QY 590 GVLITNNHVITAGATDINAFSVGSGQTGVGVYDRDQDVAVLQRLGAGLPSAIGGV 649

A;Molecule type: DNA

A;Residues: 1-393 <COL>

A;Cross-references: GB:295390; GB:AL123456; NID:g3261766; PIDN:CAB08702.1; PID:e316074; I

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: PPE

Query Match 42.9%; Score 1583; DB 2; Length 393;

Best Local Similarity 81.5%; Pred. No. 9.7e-74;

Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 142 MVDGALPPPEINSGARMYVAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSMIG 201

Db 1 MVDGALPPPEINSGARMYVAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSMIG 60

QY 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261

Db 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMT 120

QY 262 LIATNLGQNTPAIAVNEABYGMWAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 321

Db 121 LIATNLGQNTPAIAVNEABYGMWAQDAAMFGYAAATATATATALLPPEEAPMTSAGG 180

QY 322 LLEQAAVEEASDTAAANQLMNNVPOALQQAQPTGCTTPSSKLGGLWKTVPSPHSPISN 381

Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQQAQPTGCTTPSSKLGGLWKTVPSPHSPISN 240

QY 382 MVSNNHNSMTNSGVSMTNTLSMLKGFAPAAQAVOTAAQNGVRMSS---LGSSL 437

Db 438 GSSGLGGVAAANLGRAASVGLSVPOQAAWAAANQAVTPAARALPLTSLTSAERGPQMLG 497

Db 498 GLPVQMGARAG--GGLSGVLRVPRPPYVMPHSPAAG 532

Db 360 GLPLGQLTNSGGGGVSNALRMPPRAYVPRVPAAG 393

RESULT 4

F70983

Probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C;Accession: F70983

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sultston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: F70983

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-355 <COL>

A;Cross-references: GB:296071; GB:AL123456; NID:g3242254; PIDN:CAB09453.1; PID:g2181967

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: pepA

C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypsin

Query Match 26.8%; Score 989; DB 2; Length 355;

Best Local Similarity 98.0%; Pred. No. 1.5e-43;

Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 530 AAGDIAPALSDQDFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIDPN 589

Db 28 APAQAAPALSDQDFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIDPN 87

QY 590 GVLITNNHVITAGATDINAFSVGSGQTGVGVYDRDQDVAVLQRLGAGLPSAIGGV 649

Db 88 GYVLTNNHVIAGATDINAFSVSGQTYGVVDVYDRTQDVAVLQIRGAGGLPSAAGGV 147
QY 650 AYGEFVPMANGSGGGTTPRAVGRVWALQGVQASDSLTAETLNGLIQFDAAIQGD 709
Db 148 AYGEFVPMANGSGGGTTPRAVGRVWALQGVQASDSLTAETLNGLIQFDAAIQGD 207
QY 710 SGGPVVNGLGQVGVGMNTAAS 729
Db 208 SGGPVVNGLGQVGVGMNTAAS 227

RESULT 5
S47170
hypothetical protein 34k - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Dec-2002
C:Accession: S47170
R:Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
submitted to the EMBL Data Library, June 1993
A:Description: Isolation and characterisation of a 34kDa protein of Mycobacterium paratu
A:Reference number: S47170
A:Accession: S47170
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-361 <CM>
A:Cross-references: EMBL:723092; NID:G505550; PIDN:CAA80638.1; PID:G505551
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 21.1%; Score 778.5; DB 2; Length 361;
Best Local Similarity 66.0%; Pred. No. 8.1e-33;
Matches 155; Conservative 28; Mismatches 37; Indels 15; Gaps 2;

QY 495 MLGGLPVGMQARAGGLSGVLVPRPVMPHSPAAAGDIAPPALSDRFPALPDP 554
Db 15 LVGLVTVVGLGLSGVG-----LAPASA--APSGALDRFADRPPLAPDP 59
QY 555 SAMVAGVQGVVNTKLVNNVAVAGTGVIVDPNGVLTNNHVIAGATDINAFVSGGQ 614
Db 60 SAMVAGVQGVVNTKLVNNVAVAGTGVIVDPNGVLTNNHVIAGATDINAFVSGGQ 119
QY 615 TYGVVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEFVPMANGSGGGTTPRAVGR 674
Db 120 TYAVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEFVPMANGSGGGTTPRAVGR 179
QY 675 VVALGQTVQASDSLTAETLNGLIQFDAAIQGDVSGGVPVNGLGQVGVGMNTAAS 729
Db 180 VVALNQSVDATDTLTGAQENLGLIQADAPIRPGSGGVPVNSAGQVIGVDTRAT 234

RESULT 6
G70929
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R:Cole, R.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Connor, R.; Davies, K.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL020201; GB:AL123456; NID:G3250699; PIDN:CAAL17711.1; PID:al25460
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 21.0%; Score 775.5; DB 2; Length 393;
Best Local Similarity 43.4%; Pred. No. 1.3e-32;

Matches 178; Conservative 65; Mismatches 130; Indels 37; Gaps 10;
QY 143 VDFGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAQSVVWGLTVGWSIGS 202
Db 1 MDFGALPPEVNSVRYMAGPGSAPVWAAASAWGLAAELSSAATGYETVITQLSSEGWLGP 60
QY 203 SAGLVNAAAPYVWMSVTAGQAEELTAAQVVRVAAAAYETAYGLTVPPVTAENRAELMIL 262
Db 61 ASAAMEAVAPYVWMSAAAQAEQATQAAAAAFEAATAATVPPPLIAARASLMQL 120
QY 263 IATNLLGONTPALJAVNEAEYGENWAOAAAAAMFGYAAATATATATLTPFEEAPEMTSAGGL 322
Db 121 ISTNVFGONTALAAEAQYGENWAOQDSAAAYAYAGSSASASA-VTPFSTFPQIANPTAQ 179
QY 323 LEQAAAVEASDTHAA--NOLMNVPOALQOLAAQPGQTTPSSKLGGLWK----- 370
Db 180 GTOAAAVATAAGTAQSTLTETMITGLPNALOSLSPILQSS-NGPLSLWMLFEGTFNFFT 238
QY 371 -----TVSPHRSPISNMVMNMMHMTNSGVSMNTLSSMLKGFAPAAAQAVCTAA 423
Db 239 SISALLTDLOPYASFFVNTGLPYFSGMGNFIQSAKTL-GLIGSNAPAAV-----AAA 292
QY 424 QNGVRAMSSLSGLSSGLGGVVAANLGRAASVGSLSVPOAWAA--ANQAVTPAARALPLT 482
Db 293 GDAAKGLPGLGMLG---GGPVAAGLGNAAASVGLSVPPVWMSGPLPGSVTPGAAPLPVS 348
QY 483 SLTSAARERGGQMLGGIPVQMGARAGGSLGVLRVPPRYVMPHSPAAAG 532
Db 349 TVSAAPAAFPGLGGLPL----AGAGGAGAGP-RYGFRFTVMARPPFAG 393

RESULT 7
A87242
probable secreted serine proteinase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C:Accession: A87242
R: Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, F.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
cam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: A87242
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <STO>
A:Cross-references: GB:AL450380; NID:gl3093863; PIDN:CAC32191.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML2659
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 20.5%; Score 754.5; DB 2; Length 354;
Best Local Similarity 73.3%; Pred. No. 1.3e-31;
Matches 151; Conservative 25; Mismatches 25; Indels 5; Gaps 2;

QY 524 VPHSPAAAGDIAPPALSDRFPALPDPSPAMVAGVQGVVNTKLVNNVAVAGAGTG 583
Db 27 VVPGS--ATPGSPSTLDRFSNPPLNPAAWVA---PQVWNISRLGVNSAVGAGTG 81
QY 584 IVIDPNGVLTNNHVIAGATDINAFVSGGQTYGVVDVYDRTQDVAVLQIRGAGGLPSA 643
Db 82 IVIDSSGGVLTNNHVISGATDISAFDVGNGKTYGVVDVYDRTQDVAVLQIRGASNLPTA 141
QY 644 AIGGVAVGEFVPMANGSGGGTTPRAVGRVWALQGVQASDSLTAETLNGLIQFD 703
Db 142 VIGDVAIGEPFVAGNITGGGGLPSVLRVGRVVALNQTVAQSEPLTGAQETLSGLIQVDA 201
QY 704 AIQFGDSGGPVVNGLGQVGVGMNTAAS 729
Db 202 PIKFGDSGGPVVNSRGQVGVGMNTAAT 227

RESULT 8

A70932
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70932
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70932
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-409 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17729.1; PID:e125461
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 20.4%; Score 753.5; DB 2; Length 409;
Best Local Similarity 41.8%; Pred. No. 1.8e-31;
Matches 184; Conservative 52; Mismatches 139; Indels 65; Gaps 12;
QY 143 VDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIGS 202
DB 1 MDFGALPPEINSGRMYAGPGSGPLAARAAWDAALAEALYSAASYSSTIEGLTVAFWMG 60
QY 203 SAGLMVAAASPYVAMSVTAQAEELTAQVRAAAAYETAYGLTVPPPVIAENRAELMIL 262
DB 61 SSIITMAAAVAPYVAMISVTAGQAEQAGQAKIAGVYETAFATAATVPPPVIAENRALLMSL 120
QY 263 IATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPMTSAGGL 322
DB 121 VATNIFQNTPAIATBAHVAEMWAQDAAMYGAGSSATA-SQLAPFPEPPTNPSAT 179
QY 323 LEQAAVEEASDTAA- - - - -NLMNVPQALQQLAQTPTTSSKLGGLWKTVPSPH 375
DB 180 AAQSAVVAQAAGAAASDITAOISLQSLTSLPSTLQSLA--TTATATSASAG--WDTVL-- 233
QY 376 RSPISNVSMNHNHMTNSGVSTNTLSMLKGFAPAA- - - - -ARQAVQTAQ 424
DB 234 - - - - -OSITTIANTGTYSIGTIGALPGGWMLTFGQILGQAQAPGVAAL 279
QY 425 NGVR- - - - -AMSSL- - - - -GSSILGS-SGLGGVVAANLGRAASVGLSVPPQAWAAANQAVTPA 475
DB 280 LGPKAAAGALSPLAPLGGVIGDITPLGGATGCIARAIIVGSLSVPGQWAEAAAPVWRAV 339
QY 476 ARALPLTSLTSA-AERPGQMLGGLPVQMGARAGGGL- - - - -SGVLVPPRPYVWPHSP 529
DB 340 ASVLPGTGAAPALAAEAPGALFQEMALSSLAGRALAGTAVRSAGAARV- - - - - 388
QY 530 AAGDIAPPALSDQRFADPPA 549
DB 389 AGGSVTEDEVASTTIIIVIPA 408

RESULT 9

B70931
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70931

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-423 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17722.1; PID:e125461
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 20.0%; Score 737; DB 2; Length 423;
Best Local Similarity 41.8%; Pred. No. 1.3e-30;
Matches 183; Conservative 50; Mismatches 125; Indels 80; Gaps 12;
QY 143 VDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIGS 202
DB 1 MDFGALPPEINSGRMYTGPQPMLAATAWDGLAVELHATAAGVASELSALT-GAWSGP 59
QY 203 SAGLMVAAASPYVAMSVTAQAEELTAQVRAAAAYETAYGLTVPPPVIAENRAELMIL 262
DB 60 SSTSMAAAAPYVAMSVATAVHAELAGQAARLAIAAYEAAFAATVPPPVIAENRAELMVL 119
QY 263 IATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPMTSAGGL 322
DB 120 IATNIFQNTPAIMTAEQYEMWAQDAAMYGAGSSATA-SMTAFTEPPTNNGQL 178
QY 323 LEQAAVEEASDTAAAN- - - - -OLMNVPOALQQLAQTPTT- - - - -OQTTP- - - - -S 362
DB 179 GAQSSAVAQTAATAAGNLQSAFFQLLSAVPRALQGLAFTASQSASATPQWTDLGNLS 238
QY 363 SKLGLWKTVPSPHRSPISNMVMNHNHMTNSGVSTNTLSMLKGFAPAAARQAVQTA 422
DB 239 TFLGG--AVTGPTFP- - - - -GVLPSPGVPLLGIOQSVL- - - - -V 271
QY 423 AQNGVRAAMSSILGS- - - - -SLGSSGLGGG--VAANLGRAASVGLSLS 460
DB 272 TONGQGVSAALLGKIGKPFITGALAPLAEFALHTPLGSEGLGGGVSAGIGRAGLVGKLS 331
QY 461 VPOAWAAANQAVTPAARALPLTSLTS- - - - -AERPGQMLGGLPVQMGARAGGGLSGVLR 517
DB 332 VPQGVTAAPIPSPFAALQATRLAARAPIAATDGAGALLGGMALSLAGRAAAGSTG- - - 388
QY 518 VPPRPYVWPHSPAAAGDIA 535
DB 389 - - - - -HPIGSAAAPAVGAAA 403

RESULT 10

H87056
PPE-family protein (imported) - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87056
R: Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Rutter, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sgares, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H87056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL450380; NID:g13093150; PIDN:CAC31563.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML1182

Query Match 19.9%; Score 734; DB 2; Length 421;
Best Local Similarity 41.0%; Pred. No. 1.8e-30;
Matches 175; Conservative 54; Mismatches 156; Indels 42; Gaps 9;
QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 201

Db 1 MFDAALSPETNSTRMVYLGSSPILTAATAAAWVVLAKELTAAAGLQSAVEAL-LTTFEG 59
Qy 202 SSAGLMVAASPVVAMSVTACQAEITAAQVRAAAAYETAYGLTVPPVIAENAEELMI 261
Db 60 ESAALAEVTPYKELMTQNAASAEITATQLTVAANAYETATMTVPPVFNRAQACL 119
Qy 262 LIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
Db 120 LIMSIFGQNSTAIAEKEAEYTEMWQDAAMTSYQASVLEAVGATKATAPPLGVNEVG 179
Qy 322 L-----LQAAAVEASDTAAANQLMNN-----VPOALQ-----LAOP 355
Db 180 LAQEVVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 239
Qy 356 TQGTTPSSKLG--LWKTVPSPHSPISNMVMANNHMTNSGVSM-TNTLS---SMLKGFAPA 418
Db 240 QTAVPDSSAAAPQLMGGAHQHLSPINDTSLMNNHAGVANAGLSLVNGWGSAMKSLAPT 299
Qy 414 ARQAVQTAONGVRAMSSLSGLS-----SSGLGGVAAANLGRAASVGLSVPOAWAAN 477
Db 300 TTK-----AAESAFKAMGSAVGTGRGLGSSGSHVTATQAGRAASIGSLRVPQTWTAS 354
Qy 470 QAVTPAARALPLTSLSAERPGQML-GGLPVGQW---GARAGGGLSGVLVPPRPVYM 525
Db 355 QPVTATRALSPARVAVATESAPLLGGGLPWMPVPGSGGTGGVNTALRLQPRAFVM 414
Qy 526 PHSPAAG 532
Db 415 PRNPAAG 421
RESULT 11
H70931
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: H70931
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-403 <COL>
A: Cross-references: GB:AL022021; GB:AL123456; NID: G3250699; PIDN: CAAL7728.1; PID: el25461
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE
Query Match 19.9%; Score 733.5; DB 2; Length 403;
Best Local Similarity 42.9%; Pred. No. 1.9e-30;
Matches 181; Conservative 62; Mismatches 124; Indels 55; Gaps 13;
Qy 143 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSIGS 202
Db 5 LDFATLPPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSIGS 64
Qy 203 SAGLMVAASPVVAMSVTACQAEITAAQVRAAAAYETAYGLTVPPVIAENAEELMI 262
Db 65 ASAMTAAAPVYVAMSVTAVRAEQAGAAEAAYEAFAATVPPVIAENAEELMI 124
Qy 263 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 322
Db 125 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 183
Qy 323 LEQAAA-----VEBASDTAAANQLMNN-----VP--QALQ--LAOPTQGTTPSSKL 365
Db 184 AQQAATAVSVTPPLATTTAAVPLLQQLSSSLIPWYSALQWLAEALLGLTPDNRMTIV 239

Qy 364 KLGLWKVTPSPHSPISNMVMANNHMTNSGVSM-TNTLS---SMLKGFAPA-AAQQA 418
Db 244 WLDKLWALLOPN-----SNFMNTIASSGLFLPSNTIAPFLGLGGVAAAADAAGDV 293
Qy 419 VQTAONGVRAMSSLSGLS-----GLGGVAAANLGRAASVGLSVPOAWAANQAVTAA 477
Db 294 LGEATSGGLG--GALVAPLGSAGLGTGAAGLGNAAATVGLTSLVPPSTAAAPLASHLGS 351
Qy 478 AL-----PLTSLTSAERPGQMLGGLPVGQWGAARAGGGLSGVLVPP-----RPPYMPHSPA 530
Db 352 ALGGTPMVAPPVAAAG-----MPCMPFGTGGGQFG-----RAVPOYGFEPNFVAPPPA 401
Qy 531 AG 532
Db 402 AG 403
RESULT 12
G70925
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70925
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: G70925
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-408 <COL>
A: Cross-references: GB:274024; GB:AL123456; NID: G3250700; PIDN: CAAG8377.1; PID: el301025;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE
Query Match 19.0%; Score 702; DB 2; Length 408;
Best Local Similarity 42.3%; Pred. No. 7.4e-29;
Matches 180; Conservative 48; Mismatches 144; Indels 54; Gaps 13;
Qy 143 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSIGS 202
Db 1 MDFGVLPPEINSARMYAGPGSGPMWAAAADWDLAELGLAAGGYRLAISLGTAYWAGP 60
Qy 203 SAGLMVAASPVVAMSVTACQAEITAAQVRAAAAYETAYGLTVPPVIAENAEELMI 262
Db 61 AAASVAAVTPYVAMSVTACQAEITAAQVRAAAAYELAFAMTVPVPPVVAANRALLVAL 120
Qy 263 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 322
Db 121 VATNPFQONTPAIAATEAQYAEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 179
Qy 323 LEQAAA-----VEBASDTAAANQLMNN-----VP--QALQ--LAOPTQGTTPSSKL 365
Db 180 AQQAATAVSVTPPLATTTAAVPLLQQLSSSLIPWYSALQWLAEALLGLTPDNRMTIV 239
Qy 366 -----GGLWKTVPSPHSPISNMVMANNHMTNSGVSM-TNTLSMLKGRAP---A 413
Db 240 RLLGISYFDEGL-----LQFEASLAQQAIPGTPGGAG--DSGSSVLDSWGPTTIFA 287
Qy 414 AARQAVQTAONGVRAMSSIGS-----SLGSSGLGGVAAANLGRAASVGLSVPOAWAAN 469
Db 288 GPRASPVAGGAGVGGVQVTPPYWYWDRESIGSVSAALKGSSAGSLSPVPPDWAARA 347
Qy 470 QAVTPAARALPLTSLTSAERPGQMLGGLPVGQWGAARAGGGLSGVLVPPRPVYM 526
Db 348 RWANPAAWELPGDDVTALRGTAENA---LLRGFPWASAGQSTGGGF--VHKYGFRLAVNQ 402
Qy 527 HSPAAG 532

Search completed: June 30, 2004, 16:54:00
Job time : 20.0527 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 11.8689 Seconds
(without alignments)
3198.204 Million cell updates/sec

Title: US-09-597-796C-12
Perfect score: 3686
Sequence: 1 MHHEHHHTAASDNFSLQGG.....SGGPWNGLQGVGMNTAAS 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1656.5	44.9	396	1 YD61 MYCTU	Q11031 mycobacteri
2	702	19.0	408	1 YS22 MYCTU	Q10813 mycobacteri
3	687.5	18.7	463	1 Y102 MYCTU	O53951 mycobacteri
4	450	12.2	678	1 YF48 MYCTU	Q10778 mycobacteri
5	445	12.1	487	1 Y442 MYCTU	P42611 mycobacteri
6	424.5	11.5	443	1 Y878 MYCTU	Q10540 mycobacteri
7	363	9.8	408	1 SRA MYCLE	Q07297 mycobacteri
8	338	9.2	434	1 YU18 MYCTU	P31500 mycobacteri
9	332.5	9.0	435	1 YU21 MYCTU	O53268 mycobacteri
10	331	9.0	463	1 Y056 MYCTU	Q10892 mycobacteri
11	262.5	7.1	458	1 YVPA BACSU	Q9R911 bacillus su
12	233.5	6.3	178	1 YV29 MYCTU	O06246 mycobacteri
13	228	6.2	458	1 HRA2 HUMAN	O43464 homo sapien
14	225.5	6.1	448	1 DEGB ARATH	Q91u10 arabidopsis
15	225.5	6.1	864	1 ELS RAT	Q93172 rattus norv
16	225	6.1	355	1 DEGS ECOLI	P31137 escherichia
17	225	6.1	489	1 DEGP CHLPR	Q926t0 chlamydia p
18	222	6.0	437	1 DEGI ARATH	O22609 arabidopsis
19	222	6.0	455	1 DEQQ ECOLI	P33099 escherichia
20	221.5	6.0	458	1 HRA2 MOUSE	Q91iy5 mus musculu
21	221	6.0	449	1 HTPA BACSU	O43358 bacillus su
22	220.5	6.0	860	1 ELS MOUSE	P54320 mus musculu
23	219.5	6.0	400	1 YVXA BACSU	P39668 bacillus su
24	219.5	6.0	513	1 DEGP BRUSU	Q44597 brucella su
25	218.5	5.9	478	1 DEGP BUCAL	P45722 buchnera ap
26	217.5	5.9	176	1 YV25 MYCTU	Q50703 mycobacteri
27	217	5.9	466	1 HTOA HAEIN	P45129 haemophilus
28	215.5	5.8	2432	1 Y43R TRV6	P18305 chilo iride
29	215	5.8	413	1 HTEA LACHE	O242h7 lactobacill
30	215	5.8	508	1 DEGP RICCN	Q92jal rickettsia
31	213	5.7	730	1 ELS HUMAN	P15502 homo sapien
32	211.5	5.7	513	1 DEGP BRUME	Q8Yg32 brucella me
33	211	5.7	475	1 DEGP SALTY	P26982 salmonella

34 210 5.7 1783 1 RAA3 CHLRE
35 209.5 5.7 747 1 ELS BOVIN
36 208.5 5.7 474 1 DEGP ECOLI
37 208.5 5.7 497 1 DEGP CHLTR
38 208 5.6 1025 1 SLAP CAUCR
39 207.5 5.6 478 1 DEGP BUCAP
40 206.5 5.6 497 1 DEGP CHLMU
41 204.5 5.5 504 1 DEGP RHIME
42 204 5.5 513 1 DEGP RICPR
43 203 5.5 2249 1 OMPA RICRI
44 201.5 5.5 503 1 DEGP BARHE
45 197 5.3 1150 1 APMU PIG

ALIGNMENTS

RESULT 1
YD61 MYCTU STANDARD; PRT; 396 AA.
ID YD61 MYCTU
AC Q11031:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1361c/MT1406.
GN Rv1361C OR MT1406 OR MTCY02B10.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OK NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teklaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RN Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.,
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains".
RN J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z75555; CAA99966.1; -.
DR EMBL; AE007013; AA045669.1; -.
DR F0741; H70741.
DR TIGR; MT1406; -.
DR TuberculList; Rv1361c; -.
DR InterPro; IPR000030; Microbac_PPE.

DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 158 159 TA -> AT (IN REF. 2).
 SQ SEQUENCE 396 AA; 40015 MW; 6AFAED7B5F68D0 CRC64;

Query Match 44.98; Score 1656.5; DB 1; Length 396;
 Best Local Similarity 85.18; Pred.No. 2.4e-74;
 Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDFFGALPPPEINSGARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGWSWG 201
 DB 1 MVDFFGALPPPEINSGARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGWSWG 60

QY 202 SSAGLMVAASPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 261
 DB 61 SSAGLMVAASPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120

QY 262 LIATNLGQNTPAIAVNEAEYGMWQADAAAFGYAAATATATATALLPFEAPEMTSAGG 321
 DB 121 LIATNLGQNTPAIAVNEAEYGMWQADAAAFGYAAATATATATALLPFEAPEMTSAGG 180

QY 322 LLBQAAVEASDTAAANQMLNVPQALQLOAQTOGTTPSSKLGGLWKTVPSPHRSPIIN 381
 DB 181 LLEQAVAEBAIDTAANQMLNVPQALQLOAQTKSIWPFDDQLSELWKAISPHLSPLSN 240

QY 382 MVSMANNHMTNSGVSMNTTLLSMLKGFAPAPAAQAVOTAAQNGVRAMSS----LGSSL 437
 DB 241 IVSMNNHVMNTNSGVSMNTTLLSMLKGFAPAPAAQAVOTAAQNGVRAMSS----LGSSL 299

QY 438 GSSGLGGVGAANTGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAABRGPGQMUG 497
 DB 300 GSSGLGAGVAANTGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAABRGPGQMUG 359

QY 498 GLPVGOMGARAG--GGLSGLVRLVPPPPYVPHSPAAAG 532
 DB 360 GLPLGQQTNSGGGFGVSNALRMPPRAYVMPVPAAG 396

RESULT 2
 YS92_MYCTU STANDARD; PRT; 408 AA.

AC Q10813;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV2892C/MT2959/MB2916c.
 GN RV2892C OR MT2959 OR MTCY274.23C OR MB2916C.
 OS Mycobacterium tuberculosis, and
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=9829598; Pubmed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 RT Nature 393:537-544(1998).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; Pubmed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeSoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.; and
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains."
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; Pubmed=12788972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grogdian S., Lacroix C., Monsempe C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC EMBL; Z74024; CAA98377.1; --
 CC EMBL; AE007119; AAK47285.1; --
 CC EMBL; BX248344; CAG96603.1; --
 CC FIC; G70925; G70925.
 CC TIGR; MT2959; --
 CC TubercuList; RV2892c; --
 CC InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 56 POTENTIAL.
 SQ SEQUENCE 408 AA; 41469 MW; 3E3D1F20D7827199 CRC64;

Query Match 19.0%; Score 702; DB 1; Length 408;
 Best Local Similarity 42.3%; Pred.No. 9.7e-28;
 Matches 180; Conservative 48; Mismatches 144; Indels 54; Gaps 13;

QY 143 VDFGALPPEINSGARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGWSWG 202
 DB 1 MDFGVLPEINSGARMYAGPGSAPMAAAAWDSLAAELGLAAGYRLAISLTCAYWAGP 60

QY 203 SAGLMVAASPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 262
 DB 61 AASWVAATPYVAWLSATAGAQAGQAGQAAAAAAYELAFAMTVPVPPVVAANRALVAL 120

QY 263 IATNLGQNTPAIAVNEAEYGMWQADAAAFGYAAATATATATALLPFEAPEMTSAGG 322
 DB 121 VATNFGQNTPAIAATEAQVAEMWQAQAAAAAYAGSAAIAT-ELTPTTAAPVTTSPAAL 179

QY 323 LEQAAA-----VEEASDTAAANQMLN-----VP--QALQQ-LAQPTQGTTFSSKL--- 365
 DB 180 AQQAAATVSSIVTPELTAATTAAPQLQLSLSLIPWYSAALQQLAENLLGLTPDNRTIV 239

QY 366 -----GGLWKTVPSPHRSPIINSGARMYAMNNHMTNSGVSMNTTLLSMLKGFAP---A 413
 DB 240 RLLGTSYFDEGL-----LQFEASLAQQAIPGTPGGAG--DSGSSVLDGSGMTTIFA 287

QY 414 AARQAVOTAQNGVRAMSSLSGS-----SLGSSGLGGVAAANLGRAASVGSLSVPOAWAAN 469
 DB 288 GPRASPSVAGGAGVGGVQTPTPTWYWDRESIGSVSAALGKSSAGSLSVDPDWAARA 347

QY 470 QAVTPAARALP---LTSLSAAERPGQMLGGLPVGQMGARAGGGLSVLRVPPRYVMP 526
 DB 348 RWANPAARWLPQDDVTALRGTAENA---LLRGFFPMASAGQSTGGGF--VHKYFRFAVMQ 402

QY 527 HSPAAG 532
 DB 403 RPPFAG 408

```
RESULT 3
YI02_MYCTU
ID YI02_MYCTU STANDARD; PRT; 463 AA.
AC O53951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1802/MT1851/Wb1830.
GN Rv1802 OR MT1851 OR MTW049.24 OR Wb1830.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jégels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.Dovis; STRAIN=AF122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AL022021; CAA17723.1; --
CC EMBL; A5007044; AAK46123.1; --
CC EMBL; BX248340; CAD94533.1; --
CC FIC; C70931; C70931.
CC TIGR; MT1851; --
CC TubercuList; Rv1802; --
CC InterPro; IPR000030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1. Complete proteome.
CC Hypothetical protein; Complete proteome.
CC CONFLICT 401 401
CC
```

```
SQ SEQUENCE 463 AA; 46021 MW; EE64828BF09FA551 CRC64;
Query March 18.7%; Score 687.5; DB 1; Length 463;
Best Local Similarity 42.8%; Pred. No. 5.7e-27;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;
QY 143 VDFGALPPEINSEARMYAGPGSASLVAAQWDSVADLPSAASAFQSVNGLTVGWSIGS 202
DB 1 MDEGVLPPPEINSEARMYAGPGSGPLAAAWDGLATELQSTADYGSVISVL7-GVMSGQ 59
QY 203 SAGIWAASAPYVAMVSVTAGQAELETAQVRVAAAVETAYGLTVPPVIAENRAELMIL 262
DB 60 SSGTMAAAAPYVAMVSVTAGQAELETAQVRVAAAVETAYGLTVPPVIAENRAELAVL 119
QY 263 IATNLLGQNTPAIVNEAEYGENWQDAAMFCYAAATATATATLTPFEAPEMPTSAGGL 322
DB 120 AATNIFQNTGATAAAAEARVAEMWQDAAMYGAGSSSVAT-QVTFFAAPPPTTNAAGL 178
QY 323 LEQAAVBEASDTAAANQLMNVTPQALQLAQTOGTPTSSKLGGLWKTYS--PHRSPI- 379
DB 179 ATQGVAVQAQVAGSAGN-ARSLVSEVLEFLA--TAGTNNYKNTVASLMNAVTVGFYASSVY 235
QY 380 -----SNMVSMMNNHSMNTNSGVSMNTLSMLKGFAPAPAAARQAVQTAQNGVRA 429
DB 236 NSMLGLGFABSKXVLPANDVISTIFGMVQFOKFFNPVTFNF-----DLIP 282
QY 430 MSSLGSSLG-----SSGLGG---GVAANLGRAASVGSLSVPOAWAANQAVTPAARALPL 481
DB 283 KSALGAGLGLRSASISGLSGSTAPASISAGASQAGSVGMSVPPPSWAAATPAIRTVAAVFSS 342
QY 482 TSLTS--AERGGPQML-----GSLPVGQNGARAGGSLGSLVLRV 518
DB 343 TGLQAVPAAAISSGSLLSQVALASVAGGALGGAARATGGLGGGRV 389
RESULT 4
YF48_MYCTU STANDARD; PRT; 678 AA.
AC Q10778;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1548C/MT1599.
GN Rv1548C OR MT1599 OR MTCY48.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jégels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
```

laboratory strains.";
 J. Bacteriol. 184:5479-5490 (2002).
 - SIMILARITY: Belongs to the mycobacterial PPE family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; Z74020; CAA98335.1; --
 EMBL; AE007026; AAK45866.1; ALT_INIT.
 PIR; A70762; A70762.
 TIGR; MT1599; --
 TubercuList; Rv1548c; --
 InterPro; IPR000030; Microbac_PPE.
 InterPro; IPR002989; Mycobac_Pentapep.
 Pfam; PF01469; Pentapeptide_2; 11.
 Pfam; PF00823; PPE; 1.
 Hypothetical protein; Transmembrane; Complete proteome.
 TRANSMEM 14 34
 TRANSEM 180 200 POTENTIAL.
 FT CONFLICT 258 258 D -> G (IN REF. 2).
 FT SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;

 Query Match 12.2%; Score 450; DB 1; Length 678;
 Best Local Similarity 26.5%; Pred. No. 3.4e-15;
 Matches 155; Conservative 70; Mismatches 209; Indels 150; Gaps 20;

 Y 143 VDFCALPPEINARMVAGPQSASLVAAQWDSVADLSASAFQSVVWGLTVGSWIGS 202
 Db 1 MNFSVLPPPEINSLMFPAGGPGFMPLAASAWTGLDGLSAAASFAVTSQLATSGWQGP 60

 Y 203 SAGLMVAASPPYVAMVSVTAGQELTAAQVRAAAAYETAYGLTVPPIAENRAELMIL 262
 Db 61 ASAAMTGVAASYARWLTTAAQAEQAAGQAQAVSAFAEALAAATVHPGAVSANRGLRSL 120
 Y 263 IATNLGQNTPATVNEAEYGEWAQDAAMFGYAAATATATATLLPPEAEPMTSAGL 322
 Db 121 VAGNLLGQNPATVAAVEYVEQWMAADVAAMGLYHGEASAVALSUTPTFPSP----- 172
 Y 323 LEQAAVEEASDTAAANQLMN--NVPOALQQLAQPTQGTTPSSKGLGKWTVPSPHRSPTS 380
 Db 173 --SAAATPGCAVIAAGFPFLDLGV-----TIGGF----- 200

 Y 381 NTVSMANHMSTNMGVSMNTLSSMLKGFAPAAQAVQVTAQNGVRAVMSLGLSGS 440
 Db 201 ---NLASGNLGLNLG-----8FNPGSANTGVSVNLGNANIGDLNLGSGNIGSY 245
 Y 441 GLGGVAAANLG-RAASVGLSVFPQAAANAQVTPAARALPLTSLTSAERGGQWLG-- 497
 Db 246 NLGGGNTGDLNPSGNTGTILN-----WGSN-----IGSVNLGGGN-LGSY 285
 Y 498 GLFVGQMG-ARAGGGLSGVLVRPVRPVPVMPHSPAAGDIAPPALSQDRFADFPALDPPEA 556
 Db 286 NLGSGNTGDTNFGGNTGTLNVG----- 308
 Y 557 MVAQVGPQVNVNITKLYGNNAVACAGTGVIDPN---GVVLTNNHVITAGATDINAFSVSG 513
 Db 309 -----GNTGNSNFGGNTGNVNFNGNTGDTNFGSGNLGSGN---IGFNGKSGHNIGFG 360
 Y 614 QTYGVVDVVDRTQDVAVLQLRAGGLPGAA--IGGGVAVGSPVWAMGNSG----GQGGT 667
 Db 361 NS--GNNGFGLTGDQI----GFGALNSGSGNLGFGNS--GNNGIFNNGNNGNMGNS 414
 Y 668 PRAVPRVVALGQTVQASDLTGABETLNLGLQFDRAIQDQSG 711
 Db 415 GNGVGALSVEFGSSAERSGFGNSGELSTGI-----GNSG 449

RESULT 5

Y442 MYCTU
 ID Y442 MYCTU STANDARD; PRT; 487 AA.
 AC P42611; OS3727;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv0442c/MT0458.
 GN Rv0442c OR MT0458 OR MT037.06C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1173;

 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Ersmann;
 RX MEDLINE=87137260; PubMed=3029018;
 RA Shinnick T.M.;
 RT "The 65-kilodalton antigen of Mycobacterium tuberculosis.";
 RL J. Bacteriol. 169:1080-1088(1987).

 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).

 [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Pleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 Deicher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).

 - SIMILARITY: Belongs to the mycobacterial PPE family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; M15467; AAA88235.1; ALT_INIT.
 EMBL; AL021932; CAA17399.1; --
 EMBL; AE006948; AAK44681.1; --
 PIR; C70830; C70830.
 TIGR; MT0458; --
 TubercuList; Rv0442c; --
 InterPro; IPR000030; Microbac_PPE.
 InterPro; IPR002989; Mycobac_Pentapep.
 Pfam; PF01469; Pentapeptide_2; 5.
 Pfam; PF00823; PPE; 1.
 Hypothetical protein; Complete proteome.
 FT CONFLICT 40 40 E -> K (IN REF. 2).
 FT CONFLICT 96 96 I -> T (IN REF. 1).
 FT CONFLICT 211 211 G -> GNNNIG (IN REF. 1).
 SQ SEQUENCE 487 AA; 47247 MW; 97234D5B316C8C7F CRC64;

 Query Match 12.1%; Score 445; DB 1; Length 487;

Best Local Similarity 31.9%; Pred. No. 4.le-15;					
Matches 135; Conservative 49; Mismatches 181; Indels 58; Gaps 12					
QY	145	FCALPPETNSARMYAGPGSASLVAQAQWDSVADSLFSAASFQSVVMGLTGVSGMIGSSA	204		
Ddb	6	FAWLPPEININSALMPAGPGGPLIAAATAWAGELAEELLASLASLGVSITSELTSGAWLGP	65		
QY	205	GIMVAAASFFYVAMMSVTAGQAEIITAACVRVAAAAYETAYGLTVPPPVIARNAEMLILIA	264		
Ddb	66	AAMVAVATQYLAWLSLTAQAQEAQAAQAMAIAATAFEAALAANTQPAVVAANRGLMQLLAA	125		
QY	265	TNLLGONTPTATVNIEASYGENWAODAAAMCYAAATAATATALLPBEAPETSAG----	320		
Ddb	126	TNWFCQNAPALMDVEAAIEYEQWALDVAAAMAGYHFDASAAVAQLAPWCQV--LRNLGIDIG	183		
QY	321	--GLLEQAAAVEEASDTAAANQLMNVPQALQLOAQPTQCTTPSSKLG-----GLW	369		
Ddb	184	KNGQINLFGNGTGSNIGNNNIGNNIGSGNTGTGNIGSGNTGSLGNLGLDGNIGFG	243		
QY	370	KTVS-----PHRSPISNWMMNHSMTNISGVSMNTLSSMLKGFAAPAARAQVQ	420		
Ddb	244	NITGSGNIGFGITGDHQMGFGFGNSGSGN-IGFGNSGTGNVGLFNS-----	287		
QY	421	TAQQNGVRAMSSLGSSILGSSGLGGGVAANIICRAASVGLSPQAWAAAAQAVTPAARALP	480		
Ddb	288	GSGNIGSGNSGLNSGITGST---INAGLG---SAGSLNT-SFWNAGNQNALGSAAGS	340		
QY	481	LTSLTSAARCGPQM-----LGGIPVGOMGARAG--GGLSGVLIR--VPRRPYVMHPSPAA	531		
Ddb	341	EAALVSSAGYATGGMSTAALSSGILASGTLGGLGHLANVLNSGLTINTVPAAPASAFV	400		
QY	532	GDI 534			
Ddb	401	GGL 403			
 RESULT 6 Y878.MYCUTU STANDARD; PRT; 443 AA. ID Y878 MYCUTU STANDARD; PRT; 443 AA.					
AC	Q10540;				
DC	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DD	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Hypothetical pPE-family protein Rv0878c/MT0901.				
DN	Rv0878c OR MT0901 OR MTCV31.06C.				
OS	Mycobacterium tuberculosis				
OC	Bacteria; Actinobacteria; Actinomycetales;				
CC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.				
NCBI_TaxID=1773;	[1]				
RN	SEQUENCE FROM N.A.				
RC	STRAIN=H37Rv;				
RP	MEDLINE=98295987; PubMed=9634230;				
RX	Cole S.T., Brosch R., Parkhill J.J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaita F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,				
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;				
RA	"Deciphering the biology of Mycobacterium tuberculosis from the				
RL	complete genome sequence."				
RL	Nature 393:537-544(1998).				

[2] SEQUENCE FROM N.A.
RN STRAIN=CDC 1551 / Oshkosh;
RC MEDLINE=12218036;
RX STRAIN=22206494; PubMed=12218036;
RA FLEISCHMANN R.D., Alland D., Eisen J.A., Carpenter L. White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Raft D., Hickey B.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gail J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;

"Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";

J.Factoriol. 184;5479-5490(2002).

-|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
-|- SIMILARITY: Belongs to the mycobacterial PPE family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to license@isb-sib.ch).

EMBL; Z73101; CAA97385.1; --
EMBL; AE006977; AAK45143.1; ALT_INIT.
PIR; C70780; C70780.
TIGR; MT0901; --
TubercuList; Rv0878C; --
InterPro; IPR000030; Microbac_PPE.
InterPro; IPR002989; Mycobac_pentapep.
Pfam; PF01469; Pentapeptide_2; 4.
Ffam; PF00823; rPE; 1.
Hypothetical protein; Transmembrane; Repeat; Complete proteome.
TRANSMEM 15 35 POTENTIAL.
TRANSMEM 38 58 POTENTIAL.
TRANSMEM 59 79 POTENTIAL.
TRANSMEM 181 201 POTENTIAL.
DOMAIN 64 73 POLY-ALA.
DOMAIN 81 115 ALA-RICH.
DOMAIN 231 270 4 x 10 AA APPROXIMATE REPEATS.
SEQUENCE 443 AA; 43592 MW; C58BEC607F0675E2 CRC64;

Query Match 11.5%; Score 424.5; DB 1; Length 443;
Best Local Similarity 31.2%; Pred. No. 3,7e-14;
Matches 125; Conservative 56; Mismatches 132; Indels 87; Gaps 14

QY 143 VDFGALPEPTINSRMVAGPGSASLVAAAQWDSVASDLFSASAATQSIVVGLTVG---SW 199
Db 1 MNFVLPPPVNSARIYAGAPFMPLAAVAWDGLAAELGWAAASFLLISGLUTAGPGSAW 60
QY 200 IGSSAGLMVAASPYVAWMSTVQAQELTAQVRVAAAAAYETAYGLTPVPPIAENRAEL 259
Db 61 QCPAAMMAAAAYPLSLWNAATAPEGGAAGAAGAAAVVERAARAATAHPALVAANRQL 120
QY 260 MILIATNLQNTPAIAVNEAEYGEWAQDAAMFGYAATAATATATATLPFEAEPM TSA 319
Db 121 LSLVLSNLFQGNLPALIAATEASYEQLEWAQDVAAMVGHGASTVASQLTFWOO----- 173
QY 320 GCLAQCAAAVEASDTAAANLMNVFOALQQLAQPQTGTTPSKLGLKWTVPHSRPSI 379
Db 174 ---LLSVLPVVTAAPAGAV-----GVFAA--LAIPALGV---ENTG-----V 208
QY 380 SNMV---SVANNHMSTMNG-----VSMTN-----TLSSMLKGFFAPAAARQAV 419
Db 209 GNFLGIGNNNVSGSGNTGDYNFGIGNGNANLGNGINMANLGSGNAGFFNFNGNDG 268
QY 420 QTAAQNGVRASSLGS-SLGS SGGGGVAA--NLGRAASVGSLSVQVAAANAQVTPAA 476
Db 269 NTNFGSGNAGFLNIGSGNEGSGNLGFNGAGDDNTG-----WGNSGD----- 309
QY 477 RALEPLTSLSAAERPGQMGLGL--PVGOMGARAGGGLSG 514
Db 310 -----TNTGFGNSGDLNTGIGSPVTQGVANSFGNTG 341

RESULT 7
SRA_MYLE
ID_SRA_MYLE STANDARD; PRT; 408 AA.
AC_Q0297;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

RESULT 7		
SRA_MYCLE	STANDARD;	PRT; 408 AA.
ID_SRA_MYCLE		
AC Q07297		
DT 01-NOV-1995	(Rel. 32, Created)	
DT 01-NOV-1995	(Rel. 32, Last sequence update)	
DT 16-OCT-2001	(Rel. 40, Last annotation update)	

[3]
RP SEQUENCE OF 160-374 FROM N.A.
RC STRAIN-Isolate 50410;
RA Paki A.H., Dale J.W.;
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: In strain Oshkosh the gene for this protein is
CC interrupted in position 307 by an IS6110 element.
CC -!- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
CC reductase.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 294, 337 and 355.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL021287; CA416103.1; -
DR EMBL; AE007129; AAK47427.1; ALT SEQ.
DR EMBL; AE007129; AAK47430.1; ALT SEQ.
DR EMBL; X59271; CA41961.1; ALT_FRAME.
DR PIR; E70857; E70857.
DR TIGR; MT3098; -
DR Tuberculist; RV3018C; -
DR InterPro; IPR00030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;

Query Match 9.2%; Score 338; DB 1; Length 434;
Best Local Similarity 27.2%; Pred. No. 6.1e-10;
Matches 123; Conservative 54; Mismatches 211; Indels 64; Gaps 12;

QY 147 ALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVGLTVGSGWIGSSAGL 206
DB 8 ASPPEVHSHALLSAGPGPSLQAAAGWSALSAAVAQELSVVAAVAGVWGQPSAEL 67

QY 207 MYAASPYVAVMSVTAGQELTAAQVRAAAAYETAYGLTVPPPIAENRAELMILLIATN 266
DB 68 FVAATVPYVAVLQASASDAAGHEAAAGYVCAEAEMPTLPPLANHLTHAVLVATN 127

QY 267 LIGQNTPAITAVNEARYGEMWAQDAAMFYAAATATATATLLPFEEAPEMTSAGLLLEQA 326
DB 128 PFGINTIPALNEADYVRMWQAATVMSAYEAVGAAALVATHTGPAFPIVVKPG----- 181

QY 327 AAVEASDTAAAN-----QLMNVNPAALQALQPTGTPSSKLGGLKWTVSFPH 375
DB 182 --ANEASNAVAATITPPFPHIIVOFLEETFAAYDOYLSALSELPA--VAVWVQLFVD 237

QY 376 ---RSPISNNVSMNNHSMNTSGVSWNTLSSMLKGFAPAAARQAQVTAQNGVRA MSS 432
DB 238 ILGFNIIGIITLASNAQLLTFEAINASVAVGLLYAIA-GVIDIVVENVIGNLFGVVP 296

QY 433 LGSSL-----GSSGLGG--GVAA--NLGRAASVGLSVPOAAALAAQVTPAARA 478
DB 297 LGGPLLIGALAAVAVPGVAGLAGVAGLAALPAPVGAAGAPALVGVVAPVSGGWSPQAR- 355

QY 479 LPLTSLTSAERGGPGQMLGSLPGVQNGARAGG--GLSGVLVRPPRPVYPVPHSPAAGDIAP 536
DB 356 -----LVSAREPAPASTSVSVLSADRGALGFVGTAG-----KESVGPQAG 397

QY 537 PALSQDRFADFALPLDPSAMVAQVGPQVNI 568
DB 398 LTVLADEFGGAPVPMPLPGSW----GPDLVGV 425

RESULT 9
YU21_MYCTU

YU21_MYCTU STANDARD; PRT; 435 AA.
AC 053268; C53269;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3021C/RV3022C/MT3106.
GN RV3021C/RV3022C OR MT3106 OR MTV012.35C/MTV012.36C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
XP [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H3RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RC MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Paterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emoliaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490 (2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 82.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL021287; CA416106.1; ALT_FRAME.
DR EMBL; AL021287; CA416107.1; ALT_FRAME.
DR EMBL; AE007129; AAK47435.1; -
DR TIGR; MT3106; -
DR Tuberculist; RV3021C; -
DR InterPro; IPR00030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 299 299 G -> A (IN REF. 2).
FT CONFLICT 317 320 LAGV -> VTGL (IN REF. 2).
FT CONFLICT 326 326 L -> V (IN REF. 2).
SQ SEQUENCE 435 AA; 42876 MW; 3B157643EAA8484A CRC64;

Query Match 9.0%; Score 332.5; DB 1; Length 435;
Best Local Similarity 25.3%; Pred. No. 1.1e-09;
Matches 123; Conservative 59; Mismatches 172; Indels 133; Gaps 14;

QY 147 ALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVGLTVGSGWIGSSAGL 206
DB 8 ASPPEVHSHALLSAGPGPSLQAAAGWSALSAAVAQELSVVAAVAGVWGQPSAEL 67

QY 207 MYAASPYVAVMSVTAGQELTAAQVRAAAAYETAYGLTVPPPIAENRAELMILLIATN 266


```

RESULT 11
YVTA_BACSU
ID YVTA_BACSU STANDARD; PRT; 458 AA.
AC Q9R311; Q95021; Q35039;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease yvta (EC 3.4.21.-).
GN YVTA OR YVTB OR BSU33000.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=20158875; PubMed=10692364;
RA Noone D., Howell A., Devine K.M.;
RT "Expression of ykda, encoding a Bacillus subtilis homologue of HtrA, is heat shock inducible and negatively autoregulated.";
RL J. Bacteriol. 182:1592-1599 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=9804033; PubMed=9934377;
RA Kunst F., Ogasawara N., Mosser J., Albertini A.M., Alloni G.,
RA Azevedo V., Bertsch M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruchet C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiesch J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinot S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigic C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan B., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solido B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandendool M., Vanier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.P., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis";
RL Nature 390:249-256 (1997).
RN [4]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20576168; PubMed=11133960;
RA Noone D., Howell A., Collier R., Devine K.M.;
RT "ykda and yvta, HtrA-like serine proteases in Bacillus subtilis, engage in negative autoregulation and reciprocal cross-regulation of ykda and yvta gene expression.";
RL J. Bacteriol. 183:654-663 (2001).
CC -!- FUNCTION: May be involved in processing, maturation, or secretion of extracellular enzymes.

```

```

CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- INDUCTION: Induced by heat shock during exponential growth and by heterologous amylases at the transition phase of the growth cycle.
CC Negatively regulates its own expression.
CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression of htrA, especially during stress conditions.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to frameshifts in positions 87 and 246 that produce two separate ORFs.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF188296; AAF03153.1; -.
CC EMBL; Z93941; CAB07968.1; ALT_FRAME.
CC EMBL; Z93941; CAB07969.1; ALT_FRAME.
CC EMBL; Z93120; CAB15290.1; ALT_FRAME.
CC Subtilisin; SGL14155; Yvta.
CC InterPro; IPR009003; Cys_ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ_1.
CC Pfam; PF00089; tryptsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SMO0228; PDZ; 1.
CC PROSITE; PS0106; PDZ; 1.
CC Hydrolase; Protease; Serine protease; Heat shock; Transmembrane; Complete proteome.
CC DOMAIN 1 71 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 72 92 POTENTIAL.
CC DOMAIN 93 458 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 356 440 PDZ.
CC ACT_SITE 187 187 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 217 217 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 298 298 CHARGE RELAY SYSTEM (POTENTIAL).
CC SEQUENCE 458 AA; 48717 MW; 77551045A865A5CD CRC64;
CC -----
CC Query Match 7.1%; Score 262.5; DB 1; Length 458;
CC Best Local Similarity 29.1%; Pred.No. 3.1e-06;
CC Matches 85; Conservative 52; Mismatches 116; Indels 39; Gaps 14;
CC -----
CC QY 459 LSVQQAANAQAVTPAARALPLTSLTSAERFGOMLGGI-PVGOMGARAGGSL-GVL 516
CC Db 38 LDAPVSYEGRQ-----ETASALEMEKQETAVKKKRAAWLSPI--LGGIGGGLMLGI- 91
CC QY 517 RVPRPYVPHSPFAAGDIAPPALSDQDFADFPALPLDPSA---MVAQVGPQVNNI-NTK 571
CC Db 92 ----APVLPDQNOATETA-SANKQVQSDNFTTAPITNASIADWEDLEPTIVGISNIQ 146
CC QY 572 LGYNAV-----GAGTGIVT--DPNGVLTNNHVIAGATDINAFSVSGQTY 616
CC Db 147 TSQNTFTGTGGSSSESGTSGVIFKQSDKAYIITNNHVEGANKITV-TLYNGETE 205
CC QY 617 GVDVVYDRTQDVAVLQLRGAGGLPSAAIGGG--VAVGEPPVAMNGSGGQGTTPRAVEGR 674
CC Db 206 TAKLVGSDTTDLAVLEISGNVKKVASFGDSSQLRTGKVIAGNPLGQFSGTVTQGI 265
CC QY 675 VVALGQTVQASDLTGAETLNGLIQDAIQPDGSGPPVNGLGQVGVGNT 726
CC Db 266 ISGLNRTIDV-DTQGTQVEM--NVLTQDAAINPGSGPLINASSGQVIGINS 314
CC -----
CC RESULT 12
CC Y129_MYCTU
CC ID Y129_MYCTU STANDARD; PRT; 178 AA.

```


RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
[6]
RP SEQUENCE OF 134-458, INTERACTION WITH BIRC4, AND MUTAGENESIS OF
ALA-134.
RX MEDLINE=21468395; PubMed=11583623;
RA Suzuki Y., Imai Y., Nakayama H., Takahashi K., Takio K., Takahashi R.;
"A serine protease, HtrA2, is released from the mitochondria and
interacts with XIAP, inducing cell death.";
Mol. Cell 8:613-621 (2001).
[7]
RN CHARACTERIZATION.
RX MEDLINE=20334437; PubMed=10873535;
RA Savopoulos J.W., Carter P.S., Turconi S., Pettman G.R., Karren E.H.,
Gray C.W., Ward R.V., Jenkins O., Creasy C.L.;
"Expression, purification, and functional analysis of the human serine
protease HtrA2.";
Protein Expr. Purif. 19:227-234 (2000).
[8]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 134-458.
RA Li W., Srinivasula S.M., Chai J., Li P., Wu J.W., Zhang Z.,
Alnemri E.S., Shi Y.;
"Structural insights into the pro-apoptotic function of mitochondrial
serine protease HtrA2/Omi.";
Nat. Struct. Biol. 9:436-441 (2002).
CC !- FUNCTION: Serine protease that shows proteolytic activity against
a nonspecific substrate beta-casein. Promotes or induces cell
death either by direct binding to and inhibition of BIRC proteins
(also called inhibitor of apoptosis proteins, IAPs), leading to an
increase in caspase activity, or by a BIRC inhibition-independent,
caspase-independent and serine protease activity-dependent
mechanism. Isoform 2 seems to be proteolytically inactive.
CC !- SUBUNIT: Interacts with Mxi2. The mature protein, but not the
precursor binds to BIRC2, BIRC3 and BIRC4/XIAP.
CC !- SUBCELLULAR LOCATION: Mitochondrial, predominantly present in the
intermembrane space. Released into the cytosol following apoptotic
stimuli, such as UV treatment, and stimulation of mitochondria
with caspase-8 truncated BID/tBID.
CC !- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=13B;
CC IsoId=O43464-1; Sequence=Displayed;
CC Name=2; Synonyms=D-Omi;
CC IsoId=O43464-2; Sequence=VSP_005359, VSP_005361;
CC Name=3; Synonyms=p7;
CC IsoId=O43464-3; Sequence=VSP_005360, VSP_005361;
CC Name=4; Synonyms=p4;
CC IsoId=O43464-4; Sequence=VSP_005362;
CC !- TISSUE SPECIFICITY: Isoform 1 is ubiquitous; isoform 2 is
expressed predominantly in the kidney, colon and thyroid.
CC !- DOMAIN: The PDZ domain mediates interaction with Mxi2.
CC !- PTM: Autophosphorylationally activated.
CC !- SIMILARITY: Belongs to peptidase family S2C.
CC !- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; AF020760; AAB94569.2; -;
DR EMBL; AF141305; AAF66596.1; -;
DR EMBL; AF141306; AAF66597.1; -;
DR EMBL; AF141307; AAF66598.1; -;
DR EMBL; AF184911; AAG13126.1; -;
DR EMBL; AC006544; -; NOT_ANNOTATED_CDS.
DR EMBL; BC000096; AAH00096.1; -;
DR PDB; 1LCY; 29-MAY-02.

DR MEROPS; S01.278; -;
DR Genew; HGNC:14348; PRSS25.
DR MIM; 606441; -; C:Endoplasmic reticulum membrane; TAS.
DR GO; GO:0005789; C:nucleus; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003754; F:serine-type endopeptidase activity; TAS.
DR GO; GO:0004252; F:serine-type endopeptidase activity; TAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR GO; GO:0006950; P:response to stress; TAS.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00595; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Zymogen; Mitochondrion;
KW Transit peptide; Transmembrane; Apoptosis; Alternative splicing;
KW 3D-structure.
KW TRANSIT 1 31 MITOCHONDRION.
FT PROPEP 32 133
FT CHAIN 134 458 SERINE PROTEASE HTRA2.
FT TRANSMEM 105 125 POTENTIAL.
FT DOMAIN 166 342 SERINE PROTEASE.
FT DOMAIN 364 445 PDZ.
FT SITE 134 137 IAP-BINDING MOTIF.
FT ACT_SITE 198 198 CHARGE RELAY SYSTEM.
FT ACT_SITE 228 228 CHARGE RELAY SYSTEM.
FT ACT_SITE 306 306 CHARGE RELAY SYSTEM.
FT VARSPLIC 238 302 Missing (in isoform 2).
FT VARSPLIC 313 313 Missing (in isoform 3).
FT VARSPLIC 313 313 L -> LARELGAVSLQ (in isoform 3).
FT VARSPLIC 372 403 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 314 458 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 314 458 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 314 458 Missing (in isoform 2 and isoform 3).
FT MUTAGEN 134 134 A->M: LOSS OF INTERACTION WITH BIRC4;
FT MUTAGEN 134 134 LOSS OF INHIBITION OF BIRC4 ACTIVITY.
FT MUTAGEN 306 306 A->M: LOSS OF INTERACTION WITH BIRC4 AND
FT SEQUENCE 458 AA; 48840 MW; CEA955A7D0DD8C0D CRC64;
S->A: LOSS OF PROTEASE ACTIVITY.
Query Match 6.2%; Score 228; DB 1; Length 458;
Best Local Similarity 28.8%; Pred. No. 0.00015;
Matches 104; Conservative 42; Mismatches 129; Indels 86; Gaps 17;
QY 411 APAARQAVCTAAGVGRAMSSGLGSLGGVAAVLG---RAASVGLSVPAQ--- 464
DB 3 APRAGR-----GAGSLRAWALG-----GIRGRRPRITPDRLALLTSGTSPRARVT 51
QY 465 -----WAAANQAV-----TPAARA-----LPLTSLTAAERPGQMLGGLPVGQ 503
DB 52 YGTFSLWRLSVGVTEPRACLTSGTPGPRAGLTAVTPTDTRTREAESNGTSRAWLAVA- 110
QY 504 MGARA-----GGGLSGVLEVPFPPVMPHSPRAAGIAPALSDQRF---ADFPALPLD 553
DB 111 LGAGAVILLWGGG-----RGPPAVLAAPVSP-----PPASPRQYNFIAD----- 152
QY 554 PSAMVAQVGPQVNI-----NTKLGYNNVAGAGTGVIDPENGVLVNNHVIAGATDINAF 608
DB 153 ---VVEKTAFAVVYIILDRHPLGREVPISNGSGFVVAADGLIVTNAHVVAADRRVRV- 208
QY 609 SVSGQTYGVVDVGVDRDTQDVAVLQLRGAGGLPSAAIG---GGVAVGPVAVAGNSGGGG 666


```
DB 209 RLLSGDTYEAUVTVADVPVADIVATLRIQTKPLPLPLGRSADVRQGFVVAM-----G 261
QY 667 TPRAVPGRWVA-LGQTVQASDSLFGAETLNGLIQFDAAIQPGDSGGVGVNGLGVVGMN 725
DB 262 SPFALQNTITSGIVSSAQRAPDGLGLQTVNVEYIQTDAADDFGNSGGPVLNLDGIVGVN 321
QY 726 T 726
DB 322 T 322

RESULT 14
DEG8_ARATH
ID DEG8_ARATH STANDARD; PRT; 448 AA.
AC Q9LU10;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease Do-like 8, chloroplast precursor (EC 3.4.21.-).
GN DEG8 OR A5G39830 OR K13H13.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RN Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 91-113 AND 345-361.
RC STRAIN=cv. Columbia;
RA Kieselbach T., Bystedt M., Schroeder W.P.;
RL Submitted (OCT-2000) to Swiss-Prot.
CC -!- FUNCTION: Probable serine protease.
CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen
(CC (Probable)).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB024023; BAA98101.1;
CC MR00PS; S01.279;
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; trypsin; 1.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS00106; PDZ; FALSE NEG.
KW Hydrolyase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
FT TRANSIT 1 ? 90
FT TRANSIT 1 ? 90
FT CHAIN 91 448 PROTEASE DO-LIKE 8.
FT DOMAIN 152 333 SERINE PROTEASE.
FT DOMAIN 336 433 PDZ.
FT ACT_SITE 171 171 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 292 292 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 448 AA; 47492 MW; A986FC1387670AFF CRG64;
```

```
Query Match 6.1%; Score 225.5; DB 1; Length 448;
Best Local Similarity 28.1%; Pred. No. 0.0002;
Matches 87; Conservative 41; Mismatches 103; Indels 79; Gaps 13;

QY 456 VGSLSVPGAWAA-----ANQAVTPAARALPLTSLTSAERPGQMLGGLPYGQMGARAGG 510
DB 41 VSSVKVTDQWKSNNLHELAVKSPSTTRILLTSLFNLCLFNPSTYLSALALGD----- 93
QY 511 GLSGVLRVPPRPYVMPHSPAAGDIAPPALSQDREADFPALFDPS-AKVAQVGPQ-----V 565
DB 94 -----PSVATVEDVSPTV-----FPAGLFTTEGRIVOLFETKNTYSV 130
QY 566 VNI-----NTKLGVNNV-----GAGTGIVDPNGVLTNNHVIAGATDIN----- 606
DB 131 VNIFFVTLRPLQKMTGVVEIPEGNGSGVVMGQGYIVTNVHVIGNALSRNPSPGDVVGRV 190
QY 607 --AFSVGSGQTYGVVDVGVYDRTQDVAVLQKAGGLPSAAI-----GGGVAVGEPVVA 657
DB 191 NILASDGVQKNEFKGLVADRAKDLAVLKV---DAPETLLKPIKVGSGSLKVGQOCLA 246
QY 658 MNSGGQGGTTPRAVGRVVALGQTVQASDSLFGAETLNGLIQFDAAIQPGDSGGPVVNG 717
DB 247 IGNPFGPDHT--LTGVVISGLNRDI---FSQTV--TIGGGIQTDAAINFGNSGGPLDS 299
QY 718 LGQVVGVNNTA 727
DB 300 KGNLIGINTA 309

RESULT 15
ELS_RAT
ID ELS_RAT STANDARD; PRT; 864 AA.
AC Q99372;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Elastin precursor (Tropoelastin) (Fragment).
GN ELN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=91104868; PubMed=1702999;
RA Pierce R.A., Deak S.B., Stollie C.A., Boyd C.D.;
RT "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";
RL Biochemistry 29:9677-9683(1990).
[2]
RP SEQUENCE OF 781-864 FROM N.A.
RX MEDLINE=88330868; PubMed=2971041;
RA Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
RT "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";
RL J. Biol. Chem. 263:13504-13507(1988).
[3]
RP SEQUENCE OF 264-533 AND 558-864 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=92241859; PubMed=1572637;
RA Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;
RT "Elements of the rat tropoelastin gene associated with alternative
splicing.";
RL Genomics 12:651-658(1992).
CC -!- FUNCTION: Major structural protein of tissues such as aorta and
nuchal ligament, which must expand rapidly and recover completely.
CC -!- SUBUNIT: The polymeric elastin chains are cross-linked together
into an extensible 3D network.
CC -!- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=8;
Comment=Experimental confirmation may be lacking for some
isoforms;
Name=1;
IsoId=Q99372-1; Sequence=Displayed;
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 53.0977 Seconds
(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796C-12

Perfect score: 3686

Sequence: 1 MHHHHTAASDNFOLSQGG.....SGGPVNLGVGVGNMTAAS 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.25.*

- 1: sp_arChaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1944	52.7	391	16 O05298	005298 mycobacteri
2	1920.5	52.1	390	16 Q7U0E9	Q7u0e9 mycobacteri
3	1779.5	48.3	336	16 Q7U071	Q7u071 mycobacteri
4	1648	44.7	393	16 Q8VIZ3	Q8vizi3 mycobacteri
5	1583	42.9	393	16 O06341	O06341 mycobacteri
6	1583	42.9	393	16 Q7TWf5	Q7twf5 mycobacteri
7	989	26.8	355	16 O07175	O07175 mycobacteri
8	989	26.8	355	16 Q7U2S9	Q7u2s9 mycobacteri
9	778.5	21.1	361	2 Q50320	Q50320 mycobacteri
10	775.5	21.0	393	16 Q7IZJ3	Q7izj3 mycobacteri
11	775.5	21.0	411	16 O53939	O53939 mycobacteri
12	758.5	20.6	410	2 Q99011	Q99011 mycobacteri
13	758.5	20.5	354	16 Q9CCY9	Q9ccy9 mycobacteri
14	753.5	20.4	409	16 O53957	O53957 mycobacteri
15	753.5	20.4	409	16 Q7TZH7	Q7tzh7 mycobacteri
16	737	20.4	423	16 O53950	O53950 mycobacteri

17	737	20.0	423	16 Q7TZI4	Q7tzi4 mycobacteri
18	736.5	20.0	399	16 Q7TZH8	Q7tzh8 mycobacteri
19	734	19.9	421	16 Q9Z5K0	Q9z5k0 mycobacteri
20	733.5	19.9	403	16 O53956	O53956 mycobacteri
21	700	19.0	413	16 O06386	O06386 mycobacteri
22	697	18.9	391	16 P96362	P96362 mycobacteri
23	697	18.9	331	16 Q7U0T5	Q7u0t5 mycobacteri
24	696	18.9	694	16 Q8VJW0	Q8vjw0 mycobacteri
25	689	18.7	468	16 O53958	O53958 mycobacteri
26	674	18.3	380	16 P95190	P95190 mycobacteri
27	674	18.3	380	16 Q7TX66	Q7tx66 mycobacteri
28	671.5	18.2	394	16 Q7TXX5	Q7txx5 mycobacteri
29	670.5	18.2	462	16 O33310	O33310 mycobacteri
30	667.5	18.1	385	16 Q7TZR7	Q7tzt7 mycobacteri
31	664.5	18.0	385	16 O33204	O33204 mycobacteri
32	663.5	18.0	385	16 Q8VJZ0	Q8vjz0 mycobacteri
33	641	17.4	363	16 O53940	O53940 mycobacteri
34	641	17.4	364	16 Q7TZJ5	Q7tzt5 mycobacteri
35	637	17.3	350	16 Q7TZJ2	Q7tzt2 mycobacteri
36	633.5	17.2	365	16 O86373	O86373 mycobacteri
37	633.5	17.2	405	16 Q8VJW5	Q8vjw5 mycobacteri
38	628.5	17.1	381	16 Q7TX67	Q7tx67 mycobacteri
39	616.5	16.7	397	2 Q9AGF0	Q9agf0 mycobacteri
40	615.5	16.7	402	16 O33312	O33312 mycobacteri
41	615.5	16.7	443	16 Q8VKL9	Q8vkl9 mycobacteri
42	613.5	16.6	382	16 Q7TXX3	Q7txx3 mycobacteri
43	609.5	16.5	443	16 Q7U242	Q7u242 mycobacteri
44	605.5	16.4	423	16 Q7U114	Q7u114 mycobacteri
45	605.5	16.4	426	16 O05907	O05907 mycobacteri

ALIGNMENTS

RESULT 1

ID	O05298	PRELIMINARY;	PRT;	391 AA.
AC	O05298;			
DT	01-JUL-1997 (TrEMBLrel. 04, Created)			
DT	01-JUL-1997 (TrEMBLrel. 04, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Hypothetical protein (PPE family protein).			
OS	RV1196 OR MTC1364.08 OR MT1234.			
GN	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,			
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RA	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RN	Nature 393:537-544(1998).			
RC	SEQUENCE FROM N.A.			
RA	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Deicher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBSJ databases.			


```
QY 322 LLEQAAVEEASDTAAANLMNNVPOALQOQAQPTGTTTSSKLGGLWKTVPSPHRSPIGN 381
Db 181 LLEQAAVEEASDTAAANLMNNVPOALQOQAQPTGTTTSSKLGGLWKTVPSPHRSPIGN 240
QY 382 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQVQTAQNGVRAVMS-----LGSSL 437
Db 241 IVSMLNNHVMNTSGVSMNTLSSMLKGFAPAAA-QAVETAQNGVRAVMS-----LGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 359
QY 498 GLPVGQMGARAG--GGLSGVLRVPPPPYVPHSPAPAG 532
Db 360 GLPLGLG--SVNAGSGINNLRVPAZAYATPRTPAAG 393
QY 498 GLPVGQMGARAG--GGLSGVLRVPPPPYVPHSPAPAG 532
Db 360 GLPLGLG--SVNAGSGINNLRVPAZAYATPRTPAAG 393

RESULT 4
Q8VIZ3 PRELIMINARY; PRT; 393 AA.
AC Q8VIZ3
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS007161; AAK47941.1; -.
DR TIGR; MT3582; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; P:peptidase activity; IEA.
DR GO; GO:0008508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000508; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE I 1; 1.
SQ SEQUENCE 393 AA; 39688 MW; 86F0B67798855511 CRC64;

Query Match 44.7%; Score 1648; DB 16; Length 393;
Best Local Similarity 84.8%; Pred. No. 3.5e-71;
Matches 335; Conservative 20; Mismatches 34; Indels 6; Gaps 3;

QY 142 MVDGALPPEINSARMYAGPGASLVAAQWDSVADLFSASAFQSVVWGLTVGSMIG 201
Db 1 MVDGALPPEINSARMYAGPGASLVAAQWDSVADLFSASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPPYVAVMSVTAQOAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
Db 61 SSAGLMVAASPPYVAVMSVTAQOAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 262 LIATNLLGNTPTAIVNEAEYEGMWAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 321
Db 121 LIATNLLGNTPTAIVNEAEYEGMWAQDAAMFGYAAATATATATALLPPEAPLITNPGG 180
QY 322 LLEQAAVEEASDTAAANLMNNVPOALQOQAQPTGTTTSSKLGGLWKTVPSPHRSPIGN 381
Db 181 LLEQAAVEEASDTAAANLMNNVPOALQOQAQPTGTTTSSKLGGLWKTVPSPHRSPIGN 240
QY 382 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQVQTAQNGVRAVMS-----LGSSL 437
Db 241 IVSMLNNHVMNTSGVSMNTLSSMLKGFAPAAA-QAVETAQNGVRAVMS-----LGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 359
```

```
Db 241 IVSMLNNHVMNTSGVSMNTLSSMLKGFAPAAA-QAVETAQNGVRAVMS-----LGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 359
QY 498 GLPVGQMGARAG--GGLSGVLRVPPPPYVPHSPAPAG 532
Db 360 GLPLGLG--SVNAGSGINNLRVPAZAYATPRTPAAG 393

RESULT 5
Q06341 PRELIMINARY; PRT; 393 AA.
AC Q06341
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR MTCY13E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=982595987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Horsby T., Jagels K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rulston S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
DR EMBL; Z95390; CAB08702.1; -.
DR PIR; C70568; C70568.
DR TubercuList; RV3478; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 42.9%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 4.4e-68;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 142 MVDGALPPEINSARMYAGPGASLVAAQWDSVADLFSASAFQSVVWGLTVGSMIG 201
Db 1 MVDGALPPEINSARMYAGPGASLVAAQWDSVADLFSASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPPYVAVMSVTAQOAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
Db 61 SSAGLMVAASPPYVAVMSVTAQOAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 262 LIATNLLGNTPTAIVNEAEYEGMWAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 321
Db 121 LIATNLLGNTPTAIVNEAEYEGMWAQDAAMFGYAAATATATATALLPPEAPLITNPGG 180
QY 322 LLEQAAVEEASDTAAANLMNNVPOALQOQAQPTGTTTSSKLGGLWKTVPSPHRSPIGN 381
Db 181 LLEQAAVEEASDTAAANLMNNVPOALQOQAQPTGTTTSSKLGGLWKTVPSPHRSPIGN 240
QY 382 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQVQTAQNGVRAVMS-----LGSSL 437
Db 241 VSSIANNHSMGTVGSMNTLSSMLKGFAPAAA-QAVETAQNGVRAVMS-----LGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 359
```

```

Db 300 GSSGLGAGVAANLGRAASVSLVPPAWAANQAVTPAARALPLTSLTSAQAQPGHMLG 359
Qy 498 GLPVQGMARAGGSLGVLRVPPRYVMPHSPPAAG 532
Db 360 GLPLGH-SVNAGSGINNLRVPAAYAIPTPPAAG 393

RESULT 6
Q7TWFS PRELIMINARY; PRT; 393 AA.
AC Q7TWFS;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE PE family protein.
GN PP260 OR M33505.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RA Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL EMBL; BX248346; CAD95692.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 42.9%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 4.4e-68;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

Qy 142 MYDFGALPPEINSARMYAGSGASLSVAAAQWMSVASDLFSAASAFQSVVGLTVGSGWIG 201
Db 1 MYDFGALPPEINSARMYAGSGASLSVAAAQWMSVASDLFSAASAFQSVVGLTVGSGWIG 60
Qy 202 SSAGLMVAASPYVWMSVTAQAGELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 261
Db 61 SSAGLMVAASPYVWMSVTAQAGELTAQVRAAAAYETAYGLTVPPVIAENRAELMT 120
Qy 262 LIATNLLGQNTPAIAVNEAEYGENWAQDAAMFGYAAATATATATLPPPEAPEMTSAGG 321
Db 121 LTATNLLGQNTPAIEANQAYSQWQDAEAMYGAAATAATATEALLPFEDAPLIITNPGG 180
Qy 322 LLEQAAVEEASDTAAANQIMNVPQALQLOAQPTQGTTSKLLGLWKTVPSPHSPISN 381
Db 181 LLEQAAVEEADTAANQIMNVPQALQLOAQPTQGTTSKLLGLWKTVPSPHSPISN 240
Qy 382 MVSNNHMSMTNGSVNTNLTSSMLKGFAPAAARQAVQTAAGVQVAMSS----LCSSIL 437
Db 241 VSSIANHMSMMGTGVNTNLTSSMLKGLAPAAA-QAVETAENGVMSSLSGLSGSSIL 299
Qy 438 GSSGLGGVAAANLGRAASVSLVPPAWAANQAVTPAARALPLTSLTSAARGPQOMLG 497
Db 300 GSSGLGAGVAANLGRAASVSLVPPAWAANQAVTPAARALPLTSLTSAQAQPGHMLG 359
Qy 498 GLPVQGMARAGGSLGVLRVPPRYVMPHSPPAAG 532
Db 360 GLPLGH-SVNAGSGINNLRVPAAYAIPTPPAAG 393

RESULT 7
Q07175 PRELIMINARY; PRT; 355 AA.
AC Q07175;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)

```

```

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Serine protease, putative).
GN PEPA OR RV0125 OR MTC1418B.07 OR MT0133.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; Z96071; CAB09453.1; -.
DR EMBL; AE006925; AAK44357.1; -.
DR PIR; F70983; F70983.
DR TIGR; MT0133; -.
DR TubercuList; RV0125; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR008256; Peptidase_S1b_V8.
DR InterPro; IPR001940; Peptidase_Sic.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; V8PROTEASE.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Hypothetical protein; Serine protease; Protease;
KW Complete proteome.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 26.8%; Score 989; DB 16; Length 355;
Best Local Similarity 98.0%; Pred. No. 8.4e-40;
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 530 AAGDIAPALSDQRFADFPALFDPSPAMVAQVQVNVNINIKLGYNNAVAGTGVIDPN 589
Db 28 APAQAAPALSDQRFADFPALFDPSPAMVAQVQVNVNINIKLGYNNAVAGTGVIDEN 87
Qy 590 GVLITNNEVIAAGTDINAFSGVSGTGYVDVYVGRDQDVAVLQIRGAGGLPSAIGGV 649
Db 88 GVLITNNEVIAAGTDINAFSGVSGTGYVDVYVGRDQDVAVLQIRGAGGLPSAIGGV 147
Qy 650 AVGEPVWAMGSGGGTTPRVPGRVVALGQTVQASDLSLTGAETLNGLIQDAAIQPD 709

```

expressed.in vivo by Mycobacterium avium subsp paratuberculosis.";

RT Microbiology 140:1977-1982(1994).
-L- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
EMBL; Z23092; CAA80638.1; --
DR PR; S4710; S4710.
DR GO; GO:0008233; F-peptidase activity; IEA.
DR GO; GO:0004295; F-trypsin activity; IEA.
DR GO; GO:0007242; Pintracellular signaling cascade; IEA.
DR GO; GO:0006508; Proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ_1.
DR Pfam; PF00689; Crypsin; 1.
DR PRINTS; PRO0834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; signal.
FT SIGNAL 1 38 POTENTIAL.
SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match 21.1%; Score 778.5; DB 2; Length 361;
Best Local Similarity 66.0%; Pred. No. 9.5e-30;
Matches 155; Conservative 28; Mismatches 37; Indels 15; Gaps 2;

QY 495 MLGGLPVGMGARAGGGLSGVLRPVPPRYWPHSPAAAGDIAPPALSODRFADFPALPLDP 554
DB :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
15 LVGVLTIVWGVLGLGSGVG-----LAPASA--APSGIALDRPADREPLAIDP 59

QY 555 SAMVAQVGPVNNTKLGHNNAVAGTGTVIDPNGVLTNNHVIAGATINAFVSQGQ 614
DB :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
60 SAMVGQVGPVNMTKFGYNNAVAGTGTVIDPNGVLTNNHVISGATISAFDVGNGQ 119

QY 615 TYGVVDVVGYDRTOOVAQLRGAGGLPSAIGGVAVGPEVMANGSGGGGTTPRAVPGR 674
DB :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
120 TVADVVDVGYDRTOOVAQLRGAGLPPTATIGEATVGEIVALGNVGGGGTENAVAGK 179

QY 675 VVALGQTVOASDLTGAEETLINGLIOPDAIQDGSDGVPVNGLVGVVGMNTAAS 729
DB :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
180 VVALNQSVAATDTLTGAENLGGIIQADAPIKPDGSGGPMVNSAGVIGVDTAAT 234

RESULT 10
Q7TZJ3 PRELIMINARY; PRT; 393 AA.
ID Q7TZJ3 AC Q7TZJ3
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PPE family protein.
DN PPE26 OR MB1817.
GC Mycobacterium bovis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RL "The complete genome sequence of Mycobacterium bovis."
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248340; CAD94520.1; --
SQ SEQUENCE 393 AA; 38572 MW; 41944B6E547ABAB0 CRC64;

Query Match 21.0%; Score 775.5; DB 16; Length 393;
Best Local Similarity 43.4%; Pred. No. 1.5e-29;
Matches 178; Conservative 65; Mismatches 130; Indels 37; Gaps 10

Db 61 PSSITMAAAVYVAMISVTAGQAEQAQAQAKIAAGVYETAFAPVPPVIEANRALLMS 120
QY 262 LIATNLGONTFAIAVNAEYGEWAQDAAMFGYAAATATATATATLLPPEEAPETMSAGG 321
Db 121 LVATNIFGONTFAIAATEAHYAEWMAQDAAMYGAGSATA-SQAFPSPPQTTFESA 179
QY 322 LLEQAAVEEASDTAAA-----NQLMNVNPOALQQLAQPTQGTTPSSKLGGLWKTVSP 374
Db 180 TAAQSAVVAQAAGAAASDDITAQLSQLLSLPTSLQSLA--TTATATSASAG--WDTVL- 234
QY 375 HRSPTSNMVMANNHSMSTNSGVMSNTLSSMLKGFAPAA-----ARQAVQFAA 423
Db 235 -----QSITILANLTGPYSIIIGLGAIPGGMWLTGQILGAQNAAPGVAA 279
QY 424 QNGVRAMSLGSLGS-----SGLGGVAAALGRAASVGSLSVFPQAAANAQAVTP 474
Db 280 LLGPRAAAGALSPLAPLRGGYIADITPLGGGATGGIAIYVGSLSVFPQAAEAAPVRA 339
QY 475 ARAALPLISLTA-AERPGQMLGLPVQMGARAGGLJ-----SGVLRVPRFVPMHS 528
Db 340 VASVLFGTGAAPALAAEPALFGEMALSSLAGRALAGTAVRSAGAAARV----- 389
QY 529 PAAGDIAPPALSQDRFADPPA 549
Db 390 -AGGSVTEVASTTIIIVIPA 409

RESULT 13

ID Q9CCY9 PRELIMINARY; PRT; 354 AA.
AC Q9CCY9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable secreted serine protease.
GN ML2659.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583926; CAC32191.1; -;
DR PIR; A87242; A87242.
DR Leproma; ML2659; -;
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001478; PDZ_1.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR008256; Peptidase_S1B_v8.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 1.

DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9EA4A CRC64;
Query Match 20.5%; Score 754.5; DB 16; Length 354;
Best Local Similarity 73.3%; Pred. No. 1.3e-28;
Matches 151; Conservative 25; Mismatches 25; Indels 5; Gaps 2;
QY 524 VMPHSPAAGDIAPPALSQDRFADPPALPLDPSAMVAQVGPQVYVNTKLGYNNAVGAGTG 583
Db 27 VVPGS--ATSPGSLTALDRFSNRPLPLNPAAMVA---PQVYVNTKLGYNNAVGAGTG 81
QY 584 IVIDPNGVLTNNHVIAGATDINAFVSGQTVGVVGVYDRTQDYAVLQLRGAGGLPSA 643
Db 82 IVIDSSGWLVTNNHVIAGATDINAFVSGQTVGVVGVYDRTQDYAVLQLRGAGGLPSA 141
QY 644 AIGGVAVGPPVAMGNSGGCGTTPRAVPGRVVALGQTVQASDLSLGAETLNGLIQFDA 703
Db 142 VIGDVAIGEPVALGNTGGGGLPSVLPGRVVALNQTQVASEPLTGAQETLSGLIQVDA 201
QY 704 AIOPGDSGGPVVNLGQVGVGMNTAAS 729
Db 202 PIKPGDSGPPVANSRGQVGVGMNTAAT 227

RESULT 14

ID O53957 PRELIMINARY; PRT; 409 AA.
AC O53957;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1808 OR MT1856.1 OR MTU049.30.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022021; CAAL17729.1; -;
DR EMBL; AE007044; AAK46129.1; ALT_INIT.
DR PIR; A70932; A70932.
DR TIGR; MT1856.1; -;
DR TubercuList; RV1808; -;
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 409 AA; 39917 MW; 1E15202BACF36379 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 74.9614 Seconds
(without alignments)

2747.774 Million cell updates/sec

Title: US-09-597-796C-12

Perfect score: 3686

Sequence: 1 MHFHHHTAASDNFQLSQGG.....SGGPVNLGLQVVGWNTAAS 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3686	100.0	729	5	Aae29708 Mycobacte
2	3686	100.0	729	5	Aae17572 Mycobacte
3	3680	99.8	729	4	Aao22142 Ra12-H9-3
4	3680	99.8	813	7	Ada26367 Mycobacte
5	3680	99.8	825	7	Ada26366 Mycobacte
6	3680	99.8	875	7	Ada26365 Mycobacte
7	3680	99.8	930	5	Aae29731 Mycobacte
8	3680	99.8	930	7	Ada26364 Mycobacte
9	3680	99.8	1016	7	Ada26370 M. bovis
10	3680	99.8	1022	7	Ada26369 Mycobacte
11	3680	99.8	1154	7	Ada26368 Mycobacte
12	3677	99.8	729	5	Aae29709 Mycobacte
13	3677	99.8	729	5	Aae17573 Mycobacte
14	3677	99.8	729	7	Ada26374 Mycobacte
15	3670	99.6	729	7	Ada26373 Mycobacte
16	3657	99.2	729	2	Aay32059 Mycobacte
17	3628.5	98.4	726	5	Aau74588 Antigenic
18	2941	79.8	596	2	Aay32070 Mycobacte
19	2941	79.8	596	5	Aae29710 Mycobacte
20	2941	79.8	596	5	Aae17574 Mycobacte
21	2941	79.8	599	5	Aau74599 Antigenic
22	2637	71.5	1010	7	Ada26356 Mycobacte
23	2631	71.4	723	7	Ada26354 Mycobacte
24	2002	54.3	788	4	Aau01903 M. tuberc
25	2001.5	54.3	744	4	Aau01902 M. tuberc

26	2001.5	54.3	815	4	AAU01904	M. tuberc
27	1967.5	53.4	600	2	AAy32068	Mycobacte
28	1967.5	53.4	600	5	AAU74597	Antigenic
29	1944	52.7	391	2	AAW32381	Mycobacte
30	1944	52.7	391	2	AAW32449	Mycobacte
31	1944	52.7	391	2	AAW64335	Mycobacte
32	1944	52.7	391	2	AAW81702	M. tuberc
33	1944	52.7	391	2	AAy04778	Mycobacte
34	1944	52.7	391	2	AAy38989	M. tuberc
35	1944	52.7	391	2	AAy39132	M. tuberc
36	1944	52.7	391	4	AAU01888	M. tuberc
37	1944	52.7	391	5	AAE29707	Mycobacte
38	1944	52.7	391	5	AAE17571	Mycobacte
39	1939	52.6	394	2	AAy04779	Mycobacte
40	1652.5	44.8	396	2	AAW64337	Mycobacte
41	1652.5	44.8	396	2	AAW81704	M. tuberc
42	1652.5	44.8	396	2	AAy38991	M. tuberc
43	1652.5	44.8	396	2	AAy39134	M. tuberc
44	1583	42.9	393	6	ABU36968	Protein e
45	1583	42.9	393	6	ABU34531	Protein e

ALIGNMENTS

RESULT 1

AAE29708
ID AAE29708 standard; protein; 729 AA.

XX AC AAE29708;

XX AC AAE29708;

DT 29-AUG-2003 (revised)

DT 27-JAN-2003 (first entry)

XX Mycobacterium sp. MTB72F fusion protein.

DE Mycobacterium sp. MTB72F fusion protein.

XX Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen; Ra35;

KW Ra12; MTB72F; fusion protein.

XX Mycobacterium sp.

OS Mycobacterium tuberculosis.

OS Chimeric.

XX WO200272792-A2.

PN 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US008223.

PF 13-MAR-2001; 2001US-0275837P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Brannon M, Guderian J;

DR WPI; 2002-759844/82.

DR N-PSDB; AAD47083.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.

PS Disclosure; Page 90-92; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention are used in methods for eliciting immune response in mammals. They are useful as vaccines to elicit protective immunity against pathogenic

CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC MTB72P fusion protein. This fusion protein comprises Ral2 and Ra35
 CC protein from M. tuberculosis and TBH9 protein from Mycobacterium sp.
 CC (Updated on 29-AUG-2003 to standardise OS field)
 XX
 SQ Sequence 729 AA;

Query Match 100.0%; Score 3686; DB 5; Length 729;
 Best Local Similarity 100.0%; Pred. No. 1.3e-233; Indels 0; Gaps 0;
 Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
 DB 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60

QY 61 NNGNGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
 DB 61 NNGNGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120

QY 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSMYAGPGSASLVAAQWDSVASDL 180
 DB 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSMYAGPGSASLVAAQWDSVASDL 180

QY 181 FSAASAFQSVWGLTVGSMWIGSSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAYE 240
 DB 181 FSAASAFQSVWGLTVGSMWIGSSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAYE 240

QY 241 TAYGLTVPPVTAENRAELMILIAITLLGONTPALVNEAEVGEWMAODRAAMGYAAT 300
 DB 241 TAYGLTVPPVTAENRAELMILIAITLLGONTPALVNEAEVGEWMAODRAAMGYAAT 300

QY 301 ATATATLPPFEPAPMTSAGGLEQAAVEASDTAAANQLMNVPAQLQAQPTQTT 360
 DB 301 ATATATLPPFEPAPMTSAGGLEQAAVEASDTAAANQLMNVPAQLQAQPTQTT 360

QY 361 PSSKLGGLWKTSPHRSPTSNMVMNNHMTNSGVSMTNTLSMLKGFAPAARQAVQ 420
 DB 361 PSSKLGGLWKTSPHRSPTSNMVMNNHMTNSGVSMTNTLSMLKGFAPAARQAVQ 420

QY 421 TAAQNGVRAMSSGLSSGLGGVAAANLGRAASVGLSVPAQAAANQAVTFAARALP 480
 DB 421 TAAQNGVRAMSSGLSSGLGGVAAANLGRAASVGLSVPAQAAANQAVTFAARALP 480

QY 481 LTSLSAABERGPGQMLGGLPVQMGARAGGGLSGVLRVPPRYPMHSPHSPAGDIAPALS 540
 DB 481 LTSLSAABERGPGQMLGGLPVQMGARAGGGLSGVLRVPPRYPMHSPHSPAGDIAPALS 540

QY 541 QDRFADFALPLDPSAMVAQVGPVNTKLGYNNAVAGTGVIPDNGVLLTNHNVIA 600
 DB 541 QDRFADFALPLDPSAMVAQVGPVNTKLGYNNAVAGTGVIPDNGVLLTNHNVIA 600

QY 601 GATDINAFSVGGQTYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
 DB 601 GATDINAFSVGGQTYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660

QY 661 SGGGGTTPRAVGRVVALGQTVQASDSITGAEETLNGLIQFDAAIQPDSGGPVNGLGQ 720
 DB 661 SGGGGTTPRAVGRVVALGQTVQASDSITGAEETLNGLIQFDAAIQPDSGGPVNGLGQ 720

QY 721 VVGMMNTAAS 729
 DB 721 VVGMMNTAAS 729

RESULT 2
 AAE17572
 ID AAE17572 standard; protein; 729 AA.
 XX
 AC
 AA17572;
 XX

DT 22-APR-2002 (first entry)
 XX Mycobacterium species MTB72F fusion protein.
 DE
 XX Fusion protein; antigen; serological sensitivity; immune response;
 XX tuberculosis; infection; vaccine; MTB72F; Ral2-TbH9-Ra35 protein.
 KW
 XX Mycobacterium sp.
 OS
 XX
 FH Key Location/Qualifiers
 FT Region 8..139
 FT /label= Ral2_protein
 FT Region 142..532
 FT /label= TbH9FL_protein
 FT Region 535..729
 FT /label= Ra35_protein
 XX
 FN MO200198460-A2.
 XX
 XX 27-DEC-2001.
 PD
 XX 20-JUN-2001; 2001WO-US019959.
 PF
 XX 20-JUN-2000; 2000US-00597796.
 PR
 XX 01-FEB-2001; 2001US-0265737P.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX
 XX Sreiky Y, Reed S, Alderson M;
 PI
 XX WPI; 2002-147798/19.
 DR
 XX N-PSDB; AAD28342.
 XX
 PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 6; Fig 5; 136pp; English.

The present invention relates to fusion proteins containing at least two
 Mycobacterium species antigens, nucleotides encoding them and
 compositions comprising such fusion proteins. The present invention
 particularly relates to nucleic acids encoding fusion proteins that
 include two or more individual M. tuberculosis antigens which increase
 the serological sensitivity of sera from individuals infected with
 tuberculosis and methods for their use in diagnosis, prevention and
 treatment of tuberculosis infection. Sequences of the invention are
 useful for eliciting an immune response in a mammal, e.g., human,
 immunised with BCG. They are useful in the diagnosis, treatment and
 prevention of Mycobacterium infection. The fusion proteins and the
 polynucleotides are useful as diagnostic tools in patients infected with
 Mycobacterium, in vitro and in vivo assays for detecting humoral
 antibodies or cell-mediated immunity against M. tuberculosis, for the
 diagnosis of an infection or monitoring of disease progression, as
 immunogens to generate or elicit a protective immune response in a
 patient and for raising anti-M. tuberculosis antibodies in a non-human
 animal. Sequences of the invention are also used as vaccines. MTB32A
 fusion proteins of the invention are useful as in vivo diagnostic agents
 for intradermal skin test. The present sequence is a DNA encoding
 Mycobacterium species MTB72F (Ral2-TbH9-Ra35) fusion protein

Sequence 729 AA;
 Query Match 100.0%; Score 3686; DB 5; Length 729;
 Best Local Similarity 100.0%; Pred. No. 1.3e-233;
 Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
 DB 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60

QY 61 NNGNGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
 DB 61 NNGNGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120

```
QY 121 TKSGGTRTGNVTLAGPPAEFVWDFGALPPEINARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSGGTRTGNVTLAGPPAEFVWDFGALPPEINARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGWSIGSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGWSIGSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILLIATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPVIAENRAELMILLIATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAVVEASDTAAANQLMNNVPQALQLAQPTQGT 360
DB 301 ATATATLLPFEAPEMTSAGGLLEQAAAVVEASDTAAANQLMNNVPQALQLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHSMNTSGVSMNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHSMNTSGVSMNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPAQWAAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPAQWAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYNPHSPAAAGDIAPPAALS 540
DB 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYNPHSPAAAGDIAPPAALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLYNNAVGAGTGVIDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLYNNAVGAGTGVIDPNGVLTNNHVA 600
QY 601 GATDINAFSVGSGQTYGVDDVVGVDYDQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVGN 660
DB 601 GATDINAFSVGSGQTYGVDDVVGVDYDQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVGN 660
QY 661 SGGSGGTPRAVGRVVALGQTVQASDSLTAEBETLNGLIQPDAAITOPGDSGPPVWNGLGQ 720
DB 661 SGGSGGTPRAVGRVVALGQTVQASDSLTAEBETLNGLIQPDAAITOPGDSGPPVWNGLGQ 720
QY 721 VVGMMNTAAS 729
DB 721 VVGMMNTAAS 729
RESULT 3
ID AAO22142
XX AAO22142 standard; protein; 729 AA.
AC AAO22142;
XX AAO22142;
DT 03-OCT-2002 (first entry)
XX
DE Ral2-H9-32A fusion protein.
XX
KW Ral2; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
KW vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
KW immunogen; cytokines.
XX
OS Unidentified.
OS Chimeric.
XX
PN WO200125401-A2.
XX
PD 12-APR-2001.
XX
PF 06-OCT-2000; 2000WO-US027652.
XX
PR 07-OCT-1999; 99US-0158585P.
XX
PA (CORI-) CORIXA CORP.
XX
```

```
PI Skeiky Y, Guderian J;
XX WPI: 2001-266299/27.
DR N-PSDB; AAL40773.
XX
PT Recombinant nucleic acid molecule for producing high yield expression of
PT desired fusion polypeptides, encodes fusion polypeptide comprising
PT Mycobacterium tuberculosis coding sequence and heterologous polypeptide.
XX
PS Disclosure; Fig 6; 39pp; English.
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide, comprising a polynucleotide sequence of Ral2, a 14
CC kDa C-terminal fragment of serine protease antigen MTB32A of
CC Mycobacterium tuberculosis, and a heterologous polynucleotide sequence.
CC The recombinant fusion nucleic acids and polypeptides are useful for
CC providing stable and high yield expression of fusion polypeptides of both
CC eukaryotic and prokaryotic origin and to encode a protein product for use
CC as an antigen for detecting serum antibodies. The presence of serum
CC antibodies to M. tuberculosis antigens in an individual indicates that
CC the individual is infected with it. The fusion polypeptides are useful as
CC sources of proteins for monitoring binding of serum antibodies to fusion
CC proteins and as an immunogen to induce and/or enhance immune responses.
CC The coding sequences can be ligated with a coding sequence of another
CC molecule such as M. tuberculosis antigens, a cytokine or an adjuvant and
CC can be used in vivo as a DNA vaccine. This sequence represents the Ral2-
CC H9-32A fusion protein
XX
SQ Sequence 729 AA;
Query Match 99.8%; Score 3680; DB 4; Length 729;
Best Local Similarity 99.9%; Pred. No. 3.1e-233;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHEHHHTAASDNFQLSQGGQGFAPITGQAWALAGQIRSGGSPTHVIGTAFGLGVD 60
DB 1 MHHEHHHTAASDNFQLSQGGQGFAPITGQAWALAGQIRSGGSPTHVIGTAFGLGVD 60
QY 61 NNGNGARVQVWGSAPAAASLGISTGVDVITAVDGPINSATAMADALNGHHPGDIVSTWQ 120
DB 61 NNGNGARVQVWGSAPAAASLGISTGVDVITAVDGPINSATAMADALNGHHPGDIVSTWQ 120
QY 121 TKSGGTRTGNVTLAGPPAEFVWDFGALPPEINARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSGGTRTGNVTLAGPPAEFVWDFGALPPEINARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGWSIGSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGWSIGSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILLIATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPVIAENRAELMILLIATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAVVEASDTAAANQLMNNVPQALQLAQPTQGT 360
DB 301 ATATATLLPFEAPEMTSAGGLLEQAAAVVEASDTAAANQLMNNVPQALQLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHSMNTSGVSMNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHSMNTSGVSMNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPAQWAAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPAQWAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYNPHSPAAAGDIAPPAALS 540
DB 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYNPHSPAAAGDIAPPAALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLYNNAVGAGTGVIDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLYNNAVGAGTGVIDPNGVLTNNHVA 600
```

QY 601 GATDINAFSVSGGQTYGVYVVDYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGPPVAVMGN 660
 DB 601 GATDINAFSVSGGQTYGVYVVDYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGPPVAVMGN 660
 QY 661 SGGGGTTPRAVPRGVAVLQGTQVQASDSLTGABETLNGLIQFDAAIOPGSGGPPVAVMGN 720
 DB 661 SGGGGTTPRAVPRGVAVLQGTQVQASDSLTGABETLNGLIQFDAAIOPGSGGPPVAVMGN 720
 QY 721 VVGMMTAAS 729
 DB 721 VVGMMTAAS 729

RESULT 4
 ADA26367
 ID ADA26367 standard; protein; 813 AA.
 AC ADA26367;
 DT 20-NOV-2003 (first entry)
 DE Mycobacterium MTB72F-DpV (fusion MTB81F) protein.
 KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculo-static; gene therapy; vaccine.
 OS Chimeric.
 OS Mycobacterium sp.
 PN WO2003070187-A2.
 XX 28-AUG-2003.
 XX 18-FEB-2003; 2003WO-US004903.
 XX 15-FEB-2002; 2002US-0357351P.
 XX (CORI-) CORIXA CORP.
 SK Skeiky Y, Guderian J, Reed S;
 DR WPI; 2003-697554/66.
 DR N-PSDB; ADA26360.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX Claim 85; Fig 15; 112pp; English.
 XX The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculo-static activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX Sequence 813 AA;
 SQ Query Match 99.8%; Score 3680; DB 7; Length 813;
 Best Local Similarity 99.9%; Pred. No. 3.6e-233;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHTHHTAASDNFQISQGGGFAIPIGQAMATAGQIRSGGSPVHIGFTAFGLGVVD 60
 DB 1 MHHTHHTAASDNFQISQGGGFAIPIGQAMATAGQIRSGGSPVHIGFTAFGLGVVD 60
 QY 61 NNGNGARVQRVVGSAAPASLGISTGDVITAVDGPAINSATAMADALNHHHPGDIVSVTWQ 120

DB 61 NNGNGARVQRVVGSAAPASLGISTGDVITAVDGPAINSATAMADALNHHHPGDIVSVTWQ 120
 QY 121 TKSGETRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGFGSASLVAAQMWDSVASDL 180
 DB 121 TKSGETRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGFGSASLVAAQMWDSVASDL 180
 QY 181 FSAASAFOSVVMGLTVGSSAGLVAAASPPVAVMNSVTAGQAELETAQVRAAAAYE 240
 DB 181 FSAASAFOSVVMGLTVGSSAGLVAAASPPVAVMNSVTAGQAELETAQVRAAAAYE 240
 QY 241 TAYGLTVPPPPVIAENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 DB 241 TAYGLTVPPPPVIAENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPFEAEPMTSAGGLLEQAAAEBEASDTAAANGLMNNVPOALQOLAQPTGTT 360
 DB 301 ATATATLLPFEAEPMTSAGGLLEQAAAEBEASDTAAANGLMNNVPOALQOLAQPTGTT 360
 QY 361 PSSKLGGLWKTVPSPHRSPI SNMVSANNNHMTNSGVSMNTLSSMLKGFAPAAAOAVQ 420
 DB 361 PSSKLGGLWKTVPSPHRSPI SNMVSANNNHMTNSGVSMNTLSSMLKGFAPAAAOAVQ 420
 QY 421 TAAQNGVRAMSSLSGSLGGCGVAANLGRAASVGSLSVPOAAWAAANOAVTPAARALP 480
 DB 421 TAAQNGVRAMSSLSGSLGGCGVAANLGRAASVGSLSVPOAAWAAANOAVTPAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGSLGVLRVPPRPVYMPHSPAAADIAPALS 540
 DB 481 LTSLSAAERGPQMLGGLPVGQMGARAGGSLGVLRVPPRPVYMPHSPAAADIAPALS 540
 QY 541 QDRPADFPALPLDPSAMVAQVPOVNAVINTKLGYNNAVAGTGTIVDPNGVLTNNHVA 600
 DB 541 QDRPADFPALPLDPSAMVAQVPOVNAVINTKLGYNNAVAGTGTIVDPNGVLTNNHVA 600
 QY 601 GATDINAFSVSGGQTYGVYVVDYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGPPVAVMGN 660
 DB 601 GATDINAFSVSGGQTYGVYVVDYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGPPVAVMGN 660
 QY 661 SGGGGTTPRAVPRGVAVLQGTQVQASDSLTGABETLNGLIQFDAAIOPGSGGPPVAVMGN 720
 DB 661 SGGGGTTPRAVPRGVAVLQGTQVQASDSLTGABETLNGLIQFDAAIOPGSGGPPVAVMGN 720
 QY 721 VVGMMTAAS 729
 DB 721 VVGMMTAAS 729

RESULT 5
 ADA26366
 ID ADA26366 standard; protein; 825 AA.
 AC ADA26366;
 DT 20-NOV-2003 (first entry)
 DE Mycobacterium MTB72F-MTI (fusion MTB83F) protein.
 KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculo-static; gene therapy; vaccine.
 OS Chimeric.
 OS Mycobacterium sp.
 PN WO2003070187-A2.
 XX 28-AUG-2003.
 XX 18-FEB-2003; 2003WO-US004903.
 XX 15-FEB-2002; 2002US-0357351P.
 XX (CORI-) CORIXA CORP.

PI Skeiky Y, Guderian J, Reed S;
XX WPI: 2003-697554/66.
DR N-PSDB; ADA26359.
XX
PT New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX
PS Claim 85; Fig 14; 112pp; English.
XX
CC The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX
SQ Sequence 825 AA;
Query Match 99.8%; Score 3680; DB 7; Length 825;
Best Local Similarity 99.9%; Pred. No. 3.7e-233;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFQSQGGGQFAIPGQAMAIAGQIRSGGSPVHIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGGQFAIPGQAMAIAGQIRSGGSPVHIGTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAAPASLGISTGDDVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQVVGSAAPASLGISTGDDVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSGGTGTGNVTLAGPPAEFVDFGALPPEINSARMYAGFGSASLVAAQWDSVASDL 180
DB 121 TKSGGTGTGNVTLAGPPAEFVDFGALPPEINSARMYAGFGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGWSIGSAGLVMAAASPYVAMSVTAGQAEELTAQAQVRAAAYE 240
DB 181 FSAASAFQSVVWGLTVGWSIGSAGLVMAAASPYVAMSVTAGQAEELTAQAQVRAAAYE 240
QY 241 TAGYLVTPPPVTAENRAELMILLIATNLGQNTPTAVNEAEYGENWAOAAMFGYAAAT 300
DB 241 TAGYLVTPPPVTAENRAELMILLIATNLGQNTPTAVNEAEYGENWAOAAMFGYAAAT 300
QY 301 ATATATLTPFEAPEMTSAGGLLEQAAAVVEEASDTAAANQLMNNVPQALQQLAQPTQGT 360
DB 301 ATATATLTPFEAPEMTSAGGLLEQAAAVVEEASDTAAANQLMNNVPQALQQLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPTISNNVSMANNHMTNSGVMTNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPTISNNVSMANNHMTNSGVMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSLGSLGSGGVAANLGRAASVGLSVPOAAWAAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSLGSLGSGGVAANLGRAASVGLSVPOAAWAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQCMGARAGGGLSGVLVPPRPVYMPHSPAAGDIAPPAALS 540
DB 481 LTSLSAAERGPQMLGGLPVQCMGARAGGGLSGVLVPPRPVYMPHSPAAGDIAPPAALS 540
QY 541 QDRFADFPALPLDPSAMVAQVPOVYNTKLGYNNAVAGTGIVDPNGVVLTNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVPOVYNTKLGYNNAVAGTGIVDPNGVVLTNHVA 600
QY 601 GATDINAFVSGGQTYGVNDVGYDRTQDVAVILQLRGAGGLPSAAIGGGVAVGEPVYAMGN 660
DB 601 GATDINAFVSGGQTYGVNDVGYDRTQDVAVILQLRGAGGLPSAAIGGGVAVGEPVYAMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQDGGSGFVNVGLGQ 720

DB 661 SGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQDGGSGFVNVGLGQ 720
QY 721 VVGWNTAAS 729
DB 721 VVGWNTAAS 729
RESULT 6
ADA26365
ID ADA26365 standard; protein; 875 AA.
XX ADA26365;
AC ADA26365;
XX
DT 20-NOV-2003 (first entry)
XX
DE Mycobacterium MTB72F-Erd1.4 (fusion MTB89F) protein.
XX
KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
OS Chimeric.
OS Mycobacterium sp.
XX
EN WO2003070187-A2.
XX
FD 28-AUG-2003.
XX
PF 18-FEB-2003; 2003WO-US004903.
XX
PR 15-FEB-2002; 2002US-0357351P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Guderian J, Reed S;
XX
XX WPI: 2003-697554/66.
DR N-PSDB; ADA26359.
XX
PT New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX
PS Claim 85; Fig 13; 112pp; English.
XX
CC The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX
SQ Sequence 875 AA;
Query Match 99.8%; Score 3680; DB 7; Length 875;
Best Local Similarity 99.9%; Pred. No. 4e-233;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFQSQGGGQFAIPGQAMAIAGQIRSGGSPVHIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGGQFAIPGQAMAIAGQIRSGGSPVHIGTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAAPASLGISTGDDVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQVVGSAAPASLGISTGDDVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSGGTGTGNVTLAGPPAEFVDFGALPPEINSARMYAGFGSASLVAAQWDSVASDL 180
DB 121 TKSGGTGTGNVTLAGPPAEFVDFGALPPEINSARMYAGFGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGWSIGSAGLVMAAASPYVAMSVTAGQAEELTAQAQVRAAAYE 240

Db 181 PSAASAFQSVVWGLTVGSSAGLWVAASAPVVAWSVTAGQAEITAAQVRVAAAAYE 240
 Qy 241 TAYGLTVPPVPIAENRAELMILITATNLGONTFAIAVNEAEYGEWMAQDAAMFGVAAAT 300
 Db 241 TAYGLTVPPVPIAENRAELMILITATNLGONTFAIAVNEAEYGEWMAQDAAMFGVAAAT 300
 Qy 301 ATATATLLPPEEAPEMTSAGLLEQAAAEASDTAAANQLMNVPAALQQAQPTQGT 360
 Db 301 ATATATLLPPEEAPEMTSAGLLEQAAAEASDTAAANQLMNVPAALQQAQPTQGT 360
 Qy 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 Qy 421 TAAQNGVRAMSSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAAANQAVTPAARALP 480
 Db 421 TAAQNGVRAMSSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAAANQAVTPAARALP 480
 Qy 481 LTSLSAERGPQMLGGLPVGQMGARAGGLSGVLRVPRPVMPSHSPAAGDIAPPALS 540
 Db 481 LTSLSAERGPQMLGGLPVGQMGARAGGLSGVLRVPRPVMPSHSPAAGDIAPPALS 540
 Qy 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVIDPENGVLTTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVIDPENGVLTTNNHVA 600
 Qy 601 GATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660
 Db 601 GATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660
 Qy 661 SGCGGTTPRAVGRVVALGTVQASDLSLTGABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
 Db 661 SGCGGTTPRAVGRVVALGTVQASDLSLTGABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
 Qy 721 VVGNTAAS 729
 Db 721 VVGNTAAS 729

RESULT 7

AAE29731

ID AAE29731 standard; protein; 930 AA.

AC AAE29731;

XX

XX 29-AUG-2003 (revised)

DT 27-JAN-2003 (first entry)

DE

Mycobacterium sp. MTB72F-Leishmania sp. MAPS (aka r95f) fusion protein.

XX

Vaccine; immunity; diagnostic agent; gene therapy; MTB72F; MAPS;

XX

fusion protein.

XX

Mycobacterium sp.

OS Leishmania sp.

OS Chimeric.

XX

XX W0200272792-A2.

XX

PD 19-SEP-2002.

XX

PF 13-MAR-2002; 2002W0-US008223.

XX

PR 13-MAR-2001; 2001US-0275837P.

XX

PA (CORI-) CORIXA CORP.

XX

PI Skeiky Y, Brannon M, Guderian J;

XX

DR WPI; 2002-759844/82.

XX

DR N-PSDB; AAD47110.

XX

PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeifF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX

PS Example 6; Page 129-132; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeifF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC Mycobacterium sp. MTB72F-Leishmania sp. thiol specific antioxidant [TSA;
 CC MAPS (aka r95f)] fusion protein. This sequence comprises Mycobacterium
 CC species MTB72F (a 72 kDa poly-protein fusion construct comprising Rail2-
 CC Tbh9-Ra35) linked to the Leishmania TSA (MAPS). (Updated on 29-AUG-2003
 CC to standardise OS field)

XX Sequence 930 AA;

Query Match 99.8%; Score 3680; DB 5; Length 930;
 Best Local Similarity 99.9%; Pred. No. 4.3e-233;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQLSGGGGGFAIPGQAWATAGQIRSGGSGPTVHIGTAFGLGVVD 60
 Db 1 MHHHHHTAASDNFQLSGGGGGFAIPGQAWATAGQIRSGGSGPTVHIGTAFGLGVVD 60
 Qy 61 NNGNGARVQRVVGSAASLIGISTGDIVITAVDGAIPNSATAMADALNGHHHPGDIVSVTWQ 120
 Db 61 NNGNGARVQRVVGSAASLIGISTGDIVITAVDGAIPNSATAMADALNGHHHPGDIVSVTWQ 120
 Qy 121 TKSGGTGTGNTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
 Db 121 TKSGGTGTGNTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
 Qy 181 PSAASAFQSVVWGLTVGSSAGLWVAASAPVVAWSVTAGQAEITAAQVRVAAAAYE 240
 Db 181 PSAASAFQSVVWGLTVGSSAGLWVAASAPVVAWSVTAGQAEITAAQVRVAAAAYE 240
 Qy 241 TAYGLTVPPVPIAENRAELMILITATNLGONTFAIAVNEAEYGEWMAQDAAMFGVAAAT 300
 Db 241 TAYGLTVPPVPIAENRAELMILITATNLGONTFAIAVNEAEYGEWMAQDAAMFGVAAAT 300
 Qy 301 ATATATLLPPEEAPEMTSAGLLEQAAAEASDTAAANQLMNVPAALQQAQPTQGT 360
 Db 301 ATATATLLPPEEAPEMTSAGLLEQAAAEASDTAAANQLMNVPAALQQAQPTQGT 360
 Qy 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 Qy 421 TAAQNGVRAMSSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAAANQAVTPAARALP 480
 Db 421 TAAQNGVRAMSSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAAANQAVTPAARALP 480
 Qy 481 LTSLSAERGPQMLGGLPVGQMGARAGGLSGVLRVPRPVMPSHSPAAGDIAPPALS 540
 Db 481 LTSLSAERGPQMLGGLPVGQMGARAGGLSGVLRVPRPVMPSHSPAAGDIAPPALS 540
 Qy 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVIDPENGVLTTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGTGIVIDPENGVLTTNNHVA 600
 Qy 601 GATDINAFSVGSGQTYGVVDVVGVDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660

Db 601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVVMAGN 660
 QY 661 SGGGGTTPRAVPGRVVAGLQTVQASDSLTAETLNGLIQFDAAIQPGSGGPPVWNLGQ 720
 Db 661 SGGGGTTPRAVPGRVVAGLQTVQASDSLTAETLNGLIQFDAAIQPGSGGPPVWNLGQ 720
 QY 721 VVGNTAAS 729
 Db 721 VVGNTAAS 729

RESULT 8
 ADA26364
 ID ADA26364 standard; protein; 930 AA.
 XX
 AC ADA26364;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Mycobacterium MTB72F-MAPS (fusion r95F) protein.
 XX
 KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.
 XX
 OS Chimeric.
 OS Mycobacterium sp.
 XX
 FN WO2003070187-A2.
 XX
 PD 28-AUG-2003.
 XX
 PF 18-FEB-2003; 2003WO-US004903.
 XX
 PR 15-FEB-2002; 2002US-0357351P.
 XX
 PA (CORI-) CORIXA CORP.
 PI Skeiky Y, Guderian J, Reed S;
 XX
 DR WPI; 2003-697554/66.
 XX
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 and MTB39 antigens, with or without the MTB85A antigen, from a
 Mycobacterium species, useful for diagnosing, preventing and/or treating
 tuberculosis infection.
 XX
 PS Disclosure; Fig 12; 112pp; English.
 XX

The invention relates to a novel nucleic acid encoding a fusion
 polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX
 SQ Sequence 930 AA;

Query Match 99.8%; Score 3680; DB 7; Length 930;
 Best Local Similarity 99.9%; Pred. No. 4.3e-233;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEHHHHHTAASDNFOLSGGQGFAPIGQMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
 Db 1 MEHHHHHTAASDNFOLSGGQGFAPIGQMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
 QY 61 NNGGARVORVVGSAAPASLSLSTGCVITAVDGPINSATAMADALNGHHPGCVISVTWQ 120
 Db 61 NNGGARVORVVGSAAPASLSLSTGCVITAVDGPINSATAMADALNGHHPGCVISVTWQ 120
 QY 121 TKSGGTRIGNVTLAEGPPAEFWDFGALPPFINSARMYAGPGSASLVAAAQWMSVASDL 180

Db 121 TKSGGTRIGNVTLAEGPPAEFWDFGALPPFINSARMYAGPGSASLVAAAQWMSVASDL 180
 QY 181 FSAASAFOSVWGLTVGWSIGSSAGLMVAASAPYVAMSVTAGOAEULTAAQVRVAAAAAYE 240
 Db 181 FSAASAFOSVWGLTVGWSIGSSAGLMVAASAPYVAMSVTAGOAEULTAAQVRVAAAAAYE 240
 QY 241 TAYGLTVPPPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 Db 241 TAYGLTVPPPIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPFEEAPMTSAGLLECAAAVEBASPTAAANQLMNNVPQALQQLAQTQGT 360
 Db 301 ATATATLLPFEEAPMTSAGLLECAAAVEBASPTAAANQLMNNVPQALQQLAQTQGT 360
 QY 361 PSSKLGLLWKTVPSPHRSPISNMVMANNHMTNKGVSMTNTLSSMLKGRAPAAARQAVQ 420
 Db 361 PSSKLGLLWKTVPSPHRSPISNMVMANNHMTNKGVSMTNTLSSMLKGRAPAAARQAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTTPAARALP 480
 Db 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTTPAARALP 480
 QY 481 LTSLSAAERPGOMLGLPVGQMGARAGGSLGVLVPPRPYPYMPHSPAAGDIAPPALS 540
 Db 481 LTSLSAAERPGOMLGLPVGQMGARAGGSLGVLVPPRPYPYMPHSPAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
 QY 601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVVMAGN 660
 Db 601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVVMAGN 660
 QY 661 SGGGGTTPRAVPGRVVAGLQTVQASDSLTAETLNGLIQFDAAIQPGSGGPPVWNLGQ 720
 Db 661 SGGGGTTPRAVPGRVVAGLQTVQASDSLTAETLNGLIQFDAAIQPGSGGPPVWNLGQ 720
 QY 721 VVGNTAAS 729
 Db 721 VVGNTAAS 729

RESULT 9
 ADA26370
 ID ADA26370 standard; protein; 1016 AA.
 XX
 AC ADA26370;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE M. bovis MTB72F and 85b complex antigen (fusion MTB103F).
 KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine; MTB72F; 85B.
 XX
 OS Chimeric.
 OS Mycobacterium bovis.
 XX
 FN WO2003070187-A2.
 XX
 PD 28-AUG-2003.
 XX
 PF 18-FEB-2003; 2003WO-US004903.
 XX
 PR 15-FEB-2002; 2002US-0357351P.
 XX
 PA (CORI-) CORIXA CORP.
 PI Skeiky Y, Guderian J, Reed S;
 XX
 DR WPI; 2003-697554/66.

DR N-PSDB; ADA26363.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX
 PS Claim 85; Fig 18; 112pp; English.
 XX
 CC The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX
 XX Sequence 1016 AA;
 SQ
 Query Match 99.8%; Score 3680; DB 7; Length 1016;
 Best Local Similarity 99.9%; Pred. No. 4.8e-233;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MHHHHHTAASDNFQLSGGGQGFAPITGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 DB 1 MHHHHHTAASDNFQLSGGGQGFAPITGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 QY 61 NNGNGARVORVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
 DB 61 NNGNGARVORVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
 QY 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMDVSVASDL 180
 DB 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMDVSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSGSAGLVAAASPVVAMSVTACQAEITAAQVRAAAAYE 240
 DB 181 FSAASAFQSVVWGLTVGSGSAGLVAAASPVVAMSVTACQAEITAAQVRAAAAYE 240
 QY 241 TAYGLTVPPVPIAENRAELMILIAITNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 DB 241 TAYGLTVPPVPIAENRAELMILIAITNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPFEAEPMTSAGLLLEQAAVFEASPTAANOLMNNVPOALQOLAQPTGTT 360
 DB 301 ATATATLLPFEAEPMTSAGLLLEQAAVFEASPTAANOLMNNVPOALQOLAQPTGTT 360
 QY 361 PSSKLGGLWKTQVSPHRSPISNNVSWANNHMTNSGVSMTNTLSSMLKGFAPAAQAQVQ 420
 DB 361 PSSKLGGLWKTQVSPHRSPISNNVSWANNHMTNSGVSMTNTLSSMLKGFAPAAQAQVQ 420
 QY 421 TAAQNGVRAMSSLSGSSGLGGVGAANLGRAASVGSLSVPQAAWAAANQAVTPAARALP 480
 DB 421 TAAQNGVRAMSSLSGSSGLGGVGAANLGRAASVGSLSVPQAAWAAANQAVTPAARALP 480
 QY 481 LTSLSAERPGQMLGGLPVGQMGARAGGLSGVLVPPRPVYMPHSPAAAGDIAPPAL 540
 DB 481 LTSLSAERPGQMLGGLPVGQMGARAGGLSGVLVPPRPVYMPHSPAAAGDIAPPAL 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGTIVDPNGVVLTNHIVIA 600
 DB 541 QDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGTIVDPNGVVLTNHIVIA 600
 QY 601 GATDINASVSGQTYGVVDVYDRTQDAVLQLRGAGGLPSAAGGVAVGEPVWANGN 660
 DB 601 GATDINASVSGQTYGVVDVYDRTQDAVLQLRGAGGLPSAAGGVAVGEPVWANGN 660
 QY 661 SGGQCGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGGPVNGLQ 720
 DB 661 SGGQCGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGGPVNGLQ 720
 QY 721 VVGMMTAAS 729

DB 721 VVGMMTAAS 729
 RESULT 10
 ADA26369
 ID ADA26369 standard; protein; 1022 AA.
 XX
 XX ADA26369;
 AC
 XX 20-NOV-2003 (first entry)
 DT
 XX Mycobacterium MTB72F-hrCC#1 (fusion MTB102tm2F) protein.
 DE
 XX fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.
 XX
 XX Chimeric.
 OS
 XX Mycobacterium sp.
 XX
 XX WO2003070187-A2.
 FN
 XX 28-AUG-2003.
 PD
 XX 18-FEB-2003; 2003WO-US004903.
 PF
 XX 15-FEB-2002; 2002US-0357351P.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky Y, Guderian J, Reed S;
 PI
 XX WPI: 2003-697554/66.
 DR
 XX N-PSDB; ADA26362.
 DR
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 PT
 XX Claim 85; Fig 17; 112pp; English.
 PS
 XX The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 CC
 XX Sequence 1022 AA;
 SQ
 Query Match 99.8%; Score 3680; DB 7; Length 1022;
 Best Local Similarity 99.9%; Pred. No. 4.8e-233;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MHHHHHTAASDNFQLSGGGQGFAPITGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 DB 1 MHHHHHTAASDNFQLSGGGQGFAPITGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 QY 61 NNGNGARVORVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
 DB 61 NNGNGARVORVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
 QY 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMDVSVASDL 180
 DB 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMDVSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSGSAGLVAAASPVVAMSVTACQAEITAAQVRAAAAYE 240
 DB 181 FSAASAFQSVVWGLTVGSGSAGLVAAASPVVAMSVTACQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPVPIAENRAELMILITATNLLGONTPTAVNEAEYGEWMAODAAAFYAAAT 300
DB 241 TAYGLTVPVPIAENRAELMILITATNLLGONTPTAVNEAEYGEWMAODAAAFYAAAT 300
QY 301 ATATATLLPFEBAPENTSAGLLEQAAVEEASDTAAANCLMNNVPOALQQLAQTQTT 360
DB 301 ATATATLLPFEBAPENTSAGLLEQAAVEEASDTAAANCLMNNVPOALQQLAQTQTT 360
QY 361 PSSKLGGLWKTSPHRSPISNMVSNNHMSMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGGLWKTSPHRSPISNMVSNNHMSMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGGLGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSLGSSGGLGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLVPPRPYVPHSPAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLVPPRPYVPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVGAGTGIVIDPNGVVLTNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVGAGTGIVIDPNGVVLTNHVA 600
QY 601 GATDINAFSVGSGQTYGVDRYDQDVAVLQLRGAGGLPSAAITGGGVAVGEPPVAMGN 660
DB 601 GATDINAFSVGSGQTYGVDRYDQDVAVLQLRGAGGLPSAAITGGGVAVGEPPVAMGN 660
QY 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGPPVNGLGQ 720
DB 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGPPVNGLGQ 720
QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729

RESULT 11

ADA26368
ID ADA26368 standard; protein; 1154 AA.

XX AC ADA26368;

XX DT 20-NOV-2003 (first entry)

XX DE Mycobacterium MTB72F-mTCC#2 (fusion MTB114F) protein.

XX KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;

XX OS tuberculosis; tuberculostatic; gene therapy; vaccine.

XX OS Chimeric.

XX OS Mycobacterium sp.

XX PN WO2003070187-A2.

XX PD 28-AUG-2003.

XX PF 18-FEB-2003; 2003WO-US004903.

XX PR 15-FEB-2002; 2002US-0357351P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Guderian J, Reed S;

XX DR WPI; 2003-697554/66.

XX DR N-PSDB; ADA26361.

XX PT New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
XX PT and MTB39 antigens, with or without the MTB85A antigen, from a
XX PT Mycobacterium species, useful for diagnosing, preventing and/or treating
XX PT tuberculosis infection.

PS Claim 85; Fig 16; 112pp; English.

XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.

XX Sequence 1154 AA;

QY Query Match 99.8%; Score 3680; DB 7; Length 1154;

DB Best Local Similarity 99.9%; Pred. No. 5.6e-233;

Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEHHHTAASDNFQISQGGGFAIPGQAWATAGQIRSGGSPTVHIGTAFGLGVVD 60

DB 1 MEHHHTAASDNFQISQGGGFAIPGQAWATAGQIRSGGSPTVHIGTAFGLGVVD 60

QY 61 NNGNGARVQVRVGSAPASLSIGTGDVITAVDGAPINSATAMADALNHHHPGDISVTWQ 120

DB 61 NNGNGARVQVRVGSAPASLSIGTGDVITAVDGAPINSATAMADALNHHHPGDISVTWQ 120

QY 121 TKSGGTGTGNVTLAEGPPAEFMDVFGALPPEINSARMYAGPGASLVAAQMDSVASDL 180

DB 121 TKSGGTGTGNVTLAEGPPAEFMDVFGALPPEINSARMYAGPGASLVAAQMDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSSAGLMVAAAAPVVAAMSVTAGQAEITAAQVRAAAAYE 240

DB 181 FSAASAFQSVVWGLTVGSSAGLMVAAAAPVVAAMSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPVPIAENRAELMILITATNLLGONTPTAVNEAEYGEWMAODAAAFYAAAT 300

DB 241 TAYGLTVPPVPIAENRAELMILITATNLLGONTPTAVNEAEYGEWMAODAAAFYAAAT 300

QY 301 ATATATLLPFEEAPEMTSAGLLEQAAVEEASDTAAANCLMNNVPOALQQLAQTQTT 360

DB 301 ATATATLLPFEEAPEMTSAGLLEQAAVEEASDTAAANCLMNNVPOALQQLAQTQTT 360

QY 361 PSSKLGGLWKTSPHRSPISNMVSNNHMSMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420

DB 361 PSSKLGGLWKTSPHRSPISNMVSNNHMSMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420

QY 421 TAAQNGVRAMSSLGSSLGSSGGLGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480

DB 421 TAAQNGVRAMSSLGSSLGSSGGLGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480

QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLVPPRPYVPHSPAAGDIAPPALS 540

DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLVPPRPYVPHSPAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVGAGTGIVIDPNGVVLTNHVA 600

DB 541 QDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVGAGTGIVIDPNGVVLTNHVA 600

QY 601 GATDINAFSVGSGQTYGVDRYDQDVAVLQLRGAGGLPSAAITGGGVAVGEPPVAMGN 660

DB 601 GATDINAFSVGSGQTYGVDRYDQDVAVLQLRGAGGLPSAAITGGGVAVGEPPVAMGN 660

QY 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGPPVNGLGQ 720

DB 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGPPVNGLGQ 720

QY 721 VVGNTAAS 729

DB 721 VVGNTAAS 729

RESULT 12

AAE29709

ID AAE29709 standard; protein; 729 AA.

XX AC AAE29709;
 XX 29-AUG-2003 (revised)
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium sp. MTB72FmutSA fusion protein.
 XX
 XX Vaccine; immunity; diagnostic agent; gene therapy; TbH9; antigen;
 KW Ra35MutSA; Ra12; MTB72MutSA; fusion protein.
 XX
 XX Mycobacterium sp.
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 FN WO200272792-A2.
 XX
 XX 19-SEP-2002.
 PD
 XX 13-MAR-2002; 2002WO-US008223.
 PF
 XX 13-MAR-2001; 2001US-0275837P.
 FR
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky Y, Brannon M, Guderian J;
 PI
 XX WPI; 2002-759844/82.
 DR N-ESDB; RAD47084.
 XX
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX
 XX Disclosure; Page 93-95; 155pp; English.
 XX
 XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC MTB72F fusion protein. This fusion protein comprises Ra35MutSA mutant
 CC protein and Ra12 protein from M. tuberculosis and TbH9 protein from
 CC Mycobacterium sp. (Updated on 29-AUG-2003 to standardise OS field)
 XX
 XX Sequence 729 AA;
 SQ
 Query Match 99.8%; Score 3677; DB 5; Length 729;
 Best Local Similarity 99.7%; Pred. No. 5e-233;
 Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MHHHHHTAASDNPOLSQGGGPAIPGQAMATAGQIRSGGSPVTHIGTAPLFGVVD 60
 Db 1 MHHHHHTAASDNPOLSQGGGPAIPGQAMATAGQIRSGGSPVTHIGTAPLFGVVD 60
 QY 61 NNGNGARVQVWGSAPASLIGISTGVDITAVDGPINSATAMADALNGHPGDIVSVTWQ 120
 Db 61 NNGNGARVQVWGSAPASLIGISTGVDITAVDGPINSATAMADALNGHPGDIVSVTWQ 120
 QY 121 TKSGETTGNVTLAGPPAFEMVDFGNLPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 Db 121 TKSGETTGNVTLAGPPAFEMVDFGNLPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSWITGSSAGLMVAAASPYVAMWSVTAGQAEILTAAQVRVAAAAYE 240

Db 181 FSAASAFQSVVWGLTVGSWITGSSAGLMVAAASPYVAMWSVTAGQAEILTAAQVRVAAAAYE 240
 QY 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIYVNEAEYGEWMAQDAAMFGYAAAT 300
 Db 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIYVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLTPFEAEPEMTSAGLLFQAAAVEASDTAAANQLMNNVPOALQOLAQPTGGTT 360
 Db 301 ATATATLTPFEAEPEMTSAGLLFQAAAVEASDTAAANQLMNNVPOALQOLAQPTGGTT 360
 QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHSMNTSGVSMNTLSSMLKGPAPAAAOAVQ 420
 Db 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHSMNTSGVSMNTLSSMLKGPAPAAAOAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANLGRASVSGSLVPOQAWAANQAVTPAARALP 480
 Db 421 TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANLGRASVSGSLVPOQAWAANQAVTPAARALP 480
 QY 481 LTSLSAERGPOMLGLPVGQMGARAGGSLGVLRVPPRPYVMPHSPAAAGDIAPPALS 540
 Db 481 LTSLSAERGPOMLGLPVGQMGARAGGSLGVLRVPPRPYVMPHSPAAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGCTGIVDPNGVYLLTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGCTGIVDPNGVYLLTNNHVA 600
 QY 601 GATDINAFSVGSGQTYGVYVVDYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVWAMGN 660
 Db 601 GATDINAFSVGSGQTYGVYVVDYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVWAMGN 660
 QY 661 SGGQGGTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPCDSSGPPVNGLGQ 720
 Db 661 SGGQGGTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPCDSSGPPVNGLGQ 720
 QY 721 VVGMMNTAAS 729
 Db 721 VVGMMNTAAS 729
 RESULT 13
 AAE17573
 ID AAE17573 standard; protein; 729 AA.
 XX
 AC AAE17573;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium species MTB72FmutSA fusion protein.
 KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB72FmutSA; Ra12-TbH9-Ra35MutSA;
 KW mutant; mutein.
 XX
 OS Mycobacterium sp.
 XX
 FH Key Location/Qualifiers
 FT Region 8..139
 FT Region /label= Ra12_protein
 FT Region 142..532
 FT Region /label= TbH9FL_protein
 FT Region 535..729
 FT Region /label= Ra35_protein
 FT Misc-difference 710
 FT /note= "Wild type Ser substituted with Ala"
 XX
 PN WO200198460-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 20-JUN-2001; 2001WO-US019959.
 XX
 PR 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0265737P.

XX (CORI-) CORIXA CORP.
XX Skeiky Y, Reed S, Alderson M;
XX WPI; 2002-147798/19.
XX N-PGDB; AAD28343.
XX Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a subject.
XX Claim 88; Fig 5; 136pp; English.
XX The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC MTB72FmutSA (Ra12-TbH9-Ra35MutSA) mutant fusion protein
XX MTB72FmutSA (Ra12-TbH9-Ra35MutSA) mutant fusion protein
XX Sequence 729 AA;
Query Match 99.8%; Score 3677; DB 5; Length 729;
Best Local Similarity 99.7%; Pred. No. 5e-233;
Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MEHHHHTAASDNFQSQGGGFAPIGQAMALAGIQRSGGSPVHIGPTAFGLGVVD 60
Db 1 MEHHHHTAASDNFQSQGGGFAPIGQAMALAGIQRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAAPASLGI-STGDVITAVDGPINSATAMADALNGHPGDVISTWQ 120
Db 61 NNGNGARVQVVGSAAPASLGI-STGDVITAVDGPINSATAMADALNGHPGDVISTWQ 120
QY 121 TKSGGTGTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDL 180
Db 121 TKSGGTGTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWIGSAGLWMAAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWIGSAGLWMAAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVPAENRAELMILIAITNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVPAENRAELMILIAITNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEAPEMTSAGGLLEQAAVEASDTAAANCLMNNVPAQLQQLAQPTQGT 360
Db 301 ATATATLLPPEAPEMTSAGGLLEQAAVEASDTAAANCLMNNVPAQLQQLAQPTQGT 360
QY 361 PSSKLGGLWKTSPHRSPISSNMVSMANNHMTNSGVSMTNTLSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTSPHRSPISSNMVSMANNHMTNSGVSMTNTLSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLSGSLGGGGVAAANIGRAASVGSLSVQAAWAAANQAVTPBARALP 480
Db 421 TAAQNGVRAMSSLSGSLGGGGVAAANIGRAASVGSLSVQAAWAAANQAVTPBARALP 480

QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRRPVPYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRRPVPYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVQVNVNTKLGYNNAVAGCTGIVIDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVQVNVNTKLGYNNAVAGCTGIVIDPNGVLTNNHVA 600
QY 601 GATDINAFSGSGQTVGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGN 660
Db 601 GATDINAFSGSGQTVGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGN 660
QY 661 SGGGGTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIOPGDSGGPVNGLGQ 720
Db 661 SGGGGTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIOPGDSGGPVNGLGQ 720
QY 721 VVGNNMTAAS 729
Db 721 VVGNNMTAAS 729
RESULT 14
ADA26374
ID ADA26374 standard; protein; 729 AA.
XX ADA26374;
AC ADA26374;
XX 20-NOV-2003 (first entry)
DT Mycobacterium MTB72FmutSA protein.
DE Mycobacterium MTB72FmutSA protein.
XX MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A; tuberculosis;
KW tuberculostatic; Gene therapy; vaccine; MTB72F; mutant; mutcin.
XX Synthetic.
OS Mycobacterium sp.
XX Key Location/Qualifiers
FT Misc-difference 710 /note= "Wild-type Ser substituted by Ala"
XX WO2003070187-A2.
PD 28-AUG-2003.
XX 18-FEB-2003; 2003WO-US004903.
XX 15-FEB-2002; 2002US-0357351P.
XX (CORI-) CORIXA CORP.
PA Skeiky Y, Guderian J, Reed S;
XX WPI; 2003-697554/66.
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
PS Disclosure; Fig 20; 112pp; English.
XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX Sequence 729 AA;
SQ

```
Query Match      99.8%; Score 3677; DB 7; Length 729;
Best Local Similarity 99.7%; Pred. No. 5e-233;
Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSGGGQFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSGGGQFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVRVVGSAAPASLGISTGDVITAVDGAIPINSATAMADALNGHHPGDVISVTWQ 120
DB 61 NNGNGARVQVRVVGSAAPASLGISTGDVITAVDGAIPINSATAMADALNGHHPGDVISVTWQ 120

QY 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180

QY 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMWSVTAGQAEITAAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMWSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPVPIAENRAELMILIAINLLQNTPAIAVNEAEYEMWAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPVPIAENRAELMILIAINLLQNTPAIAVNEAEYEMWAQDAAMFGYAAAT 300

QY 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANOLMNNVPOALQOLAQPTGTT 360
DB 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANOLMNNVPOALQOLAQPTGTT 360

QY 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANOLMNNVPOALQOLAQPTGTT 360
DB 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANOLMNNVPOALQOLAQPTGTT 360

QY 361 PSSKLGGLWKTVPSPHRSPISNMVSMMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMVSMMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420

QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALP 480

QY 481 LTSLSAARPGQMLGGLPVGQMGARAGGLSGVLPRPVPVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAARPGQMLGGLPVGQMGARAGGLSGVLPRPVPVMPHSPAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVA 600

QY 601 GATDINAFSVSGQTYGVVDVVDRTQDVAVLQIRGAGGLPSAAIGGVAVAGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVVDVVDRTQDVAVLQIRGAGGLPSAAIGGVAVAGEPVVMGN 660

QY 661 SGGGGTTPRAVPGRVVAGTQVQASDSLTGAETLNLGFIQFADAAIQGDSGGPVVNLGQ 720
DB 661 SGGGGTTPRAVPGRVVAGTQVQASDSLTGAETLNLGFIQFADAAIQGDSGGPVVNLGQ 720

QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729

RESULT 15
ADA26373
ID ADA26373 standard; protein; 729 AA.
XX
AC ADA26373;
XX
DT 20-NOV-2003 (first entry)
XX
DE Mycobacterium wild type MBT72F protein.
XX
KW fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
XX tuberculosis; tuberculostatic; gene therapy; vaccine; MTB72F.
OS Mycobacterium sp.
XX
XX W02003070187-A2.
```

```
XX 28-AUG-2003.
XX 18-FEB-2003; 2003WO-US004903.
XX 15-FEB-2002; 2002US-0357351P.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Guderian J, Reed S;
XX WPI; 2003-697554/66.
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
XX and MTB39 antigens, with or without the MTB85A antigen, from a
XX Mycobacterium species, useful for diagnosing, preventing and/or treating
XX tuberculosis infection.
XX Disclosure; Fig 20; 112pp; English.
XX The invention relates to a novel nucleic acid encoding a fusion
XX polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
XX MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
XX A polypeptide of the invention has tuberculostatic activity. A
XX polynucleotide of the invention may have a use in gene therapy, and as a
XX vaccine. The methods and compositions of the present invention are useful
XX for diagnosing, preventing and/or treating tuberculosis infection. The
XX present sequence is used in the exemplification of the invention.
XX Sequence 729 AA;

Query Match      99.6%; Score 3670; DB 7; Length 729;
Best Local Similarity 99.7%; Pred. No. 1.4e-232;
Matches 727; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSGGGQFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSGGGQFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVRVVGSAAPASLGISTGDVITAVDGAIPINSATAMADALNGHHPGDVISVTWQ 120
DB 61 NNGNGARVQVRVVGSAAPASLGISTGDVITAVDGAIPINSATAMADALNGHHPGDVISVTWQ 120

QY 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGETRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMWSVTAGQAEITAAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMWSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPVPIAENRAELMILIAINLLQNTPAIAVNEAEYEMWAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPVPIAENRAELMILIAINLLQNTPAIAVNEAEYEMWAQDAAMFGYAAAT 300

QY 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANOLMNNVPOALQOLAQPTGTT 360
DB 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANOLMNNVPOALQOLAQPTGTT 360

QY 361 PSSKLGGLWKTVPSPHRSPISNMVSMMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMVSMMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420

QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALP 480

QY 481 LTSLSAARPGQMLGGLPVGQMGARAGGLSGVLPRPVPVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAARPGQMLGGLPVGQMGARAGGLSGVLPRPVPVMPHSPAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVA 600
```

Db	541	QDEFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGGTGIVIDPNGVVLNNHVIA	600
Qy	601	GATDINAFVSGQTYGVDVWGYDRTQDVAVLQLRGAGGLPSAAICGGYAVGEPVWAMGN	660
Db	601	GATDINAFVSGQTYGVDVWGYDRTQDVAVLQLRGAGGLPSAAICGGYAVGEPVWAMGN	660
Qy	661	SGGQGGTPRAVPGRVVVALGQTVQASDSLTCAGETLNGLIQFDAAIQPGDGGPVVWNLGQ	720
Db	661	SGGQGGTPRAVPGRVVVALGQTVQASDSLTCAGETLNGLIQFDAAIQPGDGGPVVWNLGQ	720
Qy	721	VGMNTAAS	729
Db	721	VGMNTAAS	729

Search completed: June 30, 2004, 16:48:51
Job time : 76.9614 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 43.4105 Seconds
(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796C-10
Perfect score: 3007
Sequence: 1 HHHHHHHVDFGALPPEIN.....SGGPVVGIGQVGVGNTAAS 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL 25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	64.8	391	005298	Q05298 mycobacteri
2	1920.5	63.9	390	16 Q7U089	Q7U089 mycobacteri
3	1779.5	59.2	396	16 Q7U071	Q7U071 mycobacteri
4	1648	54.8	393	16 Q8V123	Q8V123 mycobacteri
5	1583	52.6	393	16 Q06341	Q06341 mycobacteri
6	1583	52.6	393	16 Q7TWFS	Q7TWFS mycobacteri
7	989	32.9	355	16 Q07175	Q07175 mycobacteri
8	989	32.9	355	16 Q7U289	Q7U289 mycobacteri
9	779.5	25.9	393	16 Q7TZJ3	Q7TZJ3 mycobacteri
10	779.5	25.9	411	16 Q53939	Q53939 mycobacteri
11	778.5	25.9	361	2 Q50320	Q50320 mycobacteri
12	756.5	25.2	410	2 Q99Q11	Q99Q11 mycobacteri
13	754.5	25.1	354	16 Q5CCY9	Q5CCY9 mycobacteri
14	753.5	25.1	409	16 Q53957	Q53957 mycobacteri
15	751.5	25.0	409	16 Q7TZH7	Q7TZH7 mycobacteri
16	740.5	24.6	399	16 Q7TZH8	Q7TZH8 mycobacteri

17	737.5	24.5	403	16	053956	053956 mycobacteri
18	737	24.5	423	16	053950	053950 mycobacteri
19	737	24.5	423	16	07TZ14	07TZ14 mycobacteri
20	731	24.3	421	16	Q9Z5K0	Q9Z5K0 mycobacteri
21	702	23.3	391	16	P96362	P96362 mycobacteri
22	702	23.3	391	16	Q7U075	Q7U075 mycobacteri
23	702	23.3	413	16	Q06386	Q06386 mycobacteri
24	695	23.1	694	16	Q8VJW0	Q8VJW0 mycobacteri
25	690	22.9	468	16	053958	053958 mycobacteri
26	675	22.4	380	16	P95190	P95190 mycobacteri
27	675	22.4	380	16	Q7TX66	Q7TX66 mycobacteri
28	670.5	22.3	385	16	Q7TZK7	Q7TZK7 mycobacteri
29	669.5	22.3	394	16	Q7TXX5	Q7TXX5 mycobacteri
30	668.5	22.2	462	16	Q33110	Q33110 mycobacteri
31	667.5	22.2	385	16	Q33204	Q33204 mycobacteri
32	666.5	22.2	385	16	Q8VJZ0	Q8VJZ0 mycobacteri
33	641	21.3	364	16	Q7TZJ5	Q7TZJ5 mycobacteri
34	637	21.2	350	16	Q7TZJ2	Q7TZJ2 mycobacteri
35	635	21.1	363	16	053940	053940 mycobacteri
36	633.5	21.1	365	16	086373	086373 mycobacteri
37	629.5	20.9	405	16	Q8VJW5	Q8VJW5 mycobacteri
38	626.5	20.8	381	16	Q7TX67	Q7TX67 mycobacteri
39	620.5	20.6	443	16	Q8VKL9	Q8VKL9 mycobacteri
40	618.5	20.6	397	2	Q9AGF0	Q9AGF0 mycobacteri
41	615.5	20.5	382	16	Q7TXX3	Q7TXX3 mycobacteri
42	615.5	20.5	402	16	Q33312	Q33312 mycobacteri
43	614.5	20.4	443	16	Q7U242	Q7U242 mycobacteri
44	606.5	20.2	423	16	Q7U114	Q7U114 mycobacteri
45	606.5	20.2	426	16	C05907	C05907 mycobacteri

ALIGNMENTS

RESULT 1
005298 PRELIMINARY; PRT; 391 AA.
ID O05298
AC O05298;
DT 01-JUL-1997 (TRENBLREL. 04, Created)
DT 01-JUL-1997 (TRENBLREL. 04, Last sequence update)
DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
DE Hypothetical protein (PPE family protein).
GN RV1196 OR MTC1364.08 OR MTL234.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teklaia F.,
Badock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
Hornsey T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence".
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Emdolaeva M.D., Salzberg S.L.,
Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bisshai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains".
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

EMBL; Z93777; CAB07839.1; -.
 EMBL; AB007000; AAK45491.1; -.
 PIR; B70608; B70608.
 TIGR; MT1234; -.
 Tuberculist; Rv1196; -.
 InterPro: IPR000030; Microbac_PPE.
 Pfam; PF00823; PPE; 1.
 Hypothetical protein; Complete proteome.
 SEQUENCE 391 AA; 39158 MW; 540939683ABDC0F8 CRC64;

Query Match 64.8%; Score 1949; DB 16; Length 391;
 Best Local Similarity 100.0%; Pred. No. 3.6e-90;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

9 MVDFGALPPEINSARMYAGPGSASLVAAQAQWMDVSADLFSASAFAQSVVWGLTVGSGWIG 68
 1 MVDFGALPPEINSARMYAGPGSASLVAAQAQWMDVSADLFSASAFAQSVVWGLTVGSGWIG 60
 69 SSAGLMVAAAAPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
 61 SSAGLMVAAAAPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
 129 LIATNLGQNTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATATLPPFEAPEMTSAGG 188
 121 LIATNLGQNTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 180
 189 LLEQAAVEEASDTAAANQLMNNVPQALQOLAQPTQGTTPSSKGLGWLKTVSPHRSPISN 248
 181 LLEQAAVEEASDTAAANQLMNNVPQALQOLAQPTQGTTPSSKGLGWLKTVSPHLSPISN 240
 249 MVSANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSGSSG 308
 241 MVSANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSGSSG 299
 309 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAAEERPGOMLGLPLV 368
 300 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAAEERPGOMLGLPLV 359
 369 QMGARAGGSLGVLRVPPRPYVMPHSPAAG 399
 360 QMGARAGGSLGVLRVPPRPYVMPHSPAAG 390

RESULT 3
 Q7U071 PRELIMINARY; PRT; 396 AA.
 AC Q7U071;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE PPE family protein.
 GN PPE19 OR MB1396C.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA "The complete genome sequence of Mycobacterium bovis."
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 RL EMBL; BX248338; CAD94089.1; -.
 KW Complete proteome.
 SQ SEQUENCE 396 AA; 39651 MW; 698A92A50E3CBA3F CRC64;

Query Match 59.2%; Score 1779.5; DB 16; Length 396;
 Best Local Similarity 91.7%; Pred. No. 1.1e-81;
 Matches 364; Conservative 9; Mismatches 17; Indels 7; Gaps 3;

9 MVDFGALPPEINSARMYAGPGSASLVAAQAQWMDVSADLFSASAFAQSVVWGLTVGSGWIG 68
 1 MVDFGALPPEINSARMYAGPGSASLVAAQAQWMDVSADLFSASAFAQSVVWGLTVGSGWIG 60
 69 SSAGLMVAAAAPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
 61 SSAGLMVAAAAPYVAMSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
 129 LIATNLGQNTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 188
 121 LIATNLGQNTPAIAVNEAEYGENWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 180
 189 LLEQAAVEEASDTAAANQLMNNVPQALQOLAQPTQGTTPSSKGLGWLKTVSPHRSPISN 248
 181 LLEQAAVEEASDTAAANQLMNNVPQALQOLAQPTQGTTPSSKGLGWLKTVSPHLSPISN 240
 249 MVSANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSGSSG 308
 241 MVSANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSGSSG 299
 309 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAAEERPGOMLGLPLV 368
 300 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAAEERPGOMLGLPLV 359
 369 QMGARAGGSLGVLRVPPRPYVMPHSPAAG 399
 360 QMGARAGGSLGVLRVPPRPYVMPHSPAAG 390

RESULT 2
 Q7U0E9 PRELIMINARY; PRT; 390 AA.
 AC Q7U0E9;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE PPE family protein.
 GN PPE18 OR MB1228.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA "The complete genome sequence of Mycobacterium bovis."
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 RL EMBL; BX248338; CAD94089.1; -.
 KW Complete proteome.
 SQ SEQUENCE 390 AA; 39013 MW; 42788276BAB0B436 CRC64;

Query Match 63.9%; Score 1920.5; DB 16; Length 390;
 Best Local Similarity 99.0%; Pred. No. 9.6e-89;
 Matches 367; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

```

QY 189 LLEQAAVEEASDTAAANOLMNNVPOALQOLAQPTGGTTPSSKLGGLWKTVPSPHSPLN 248
DB 181 LLEQAAVEEASDTAAANOLMNNVPOALQOLAQPTGGTTPSSKLGGLWKTVPSPHSPLN 240
QY 249 MYSMANNHMTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSS-----LGSSL 304
DB 241 IVSMLNNHVSMTSGVSMNTLSSMLKGFAP-AAAQAVETAQAQNGVQAMSSLSGSLGSSL 299
QY 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMLG 364
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPPAWAANAQVTPAARALPLTSLTSAQAQTPAGHMLG 359
QY 365 GLPVGMGARAG--GGLSGVLVRPPVPMPSHAAG 399
DB 360 GLPLGLH-SYNAGSINNLRVPAVAIPIPTPAAG 393

RESULT 4
Q8VIZ3 PRELIMINARY; PRT; 393 AA.
AC Q8VIZ3
DT 01-MAR-2002 (TREMREL. 20, Created)
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)
DE PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Felschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Bishai A., Ueberback T., Weidman J., Kfoury H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007161; AAK47941.1; -
DR TIGR; MT3582; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR000508; Peptidase_S26.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE_1; 1.
SQ SEQUENCE 393 AA; 39688 MW; 86F0B67798855511 CRC64;

Query Match 54.8%; Score 1648; DB 16; Length 393;
Best Local Similarity 84.8%; Pred. No. 4.3e-75;
Matches 335; Conservative 20; Mismatches 34; Indels 6; Gaps 3;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 69 SSAGLWVAASPYVAVMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
DB 61 SSAGLWVAASPYVAVMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
QY 129 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPMTSAGG 188
DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATEALLPFEAPLITNPGG 180
QY 189 LLEQAAVEEASDTAAANOLMNNVPOALQOLAQPTGGTTPSSKLGGLWKTVPSPHSPLN 248
DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATEALLPFEAPLITNPGG 180
QY 189 LLEQAAVEEASDTAAANOLMNNVPOALQOLAQPTGGTTPSSKLGGLWKTVPSPHSPLN 248
DB 181 LLEQAAVEEASDTAAANOLMNNVPOALQOLAQPTGGTTPSSKLGGLWKTVPSPHSPLN 240
QY 249 MYSMANNHMTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSS-----LGSSL 304
DB 181 LLEQAAVEEASDTAAANOLMNNVPOALQOLAQPTGGTTPSSKLGGLWKTVPSPHSPLN 240
QY 249 MYSMANNHMTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSS-----LGSSL 304
DB 241 LLEQAAVEEASDTAAANOLMNNVPOALQOLAQPTGGTTPSSKLGGLWKTVPSPHSPLN 240
QY 249 MYSMANNHMTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSS-----LGSSL 304

```

```

DB 241 IVSMLNNHVSMTSGVSMNTLSSMLKGFAP-AAAQAVETAQAQNGVQAMSSLSGSLGSSL 299
QY 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMLG 364
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAQAQTPAGHMLG 359
QY 365 GLPVGMGARAGGGLSGVLVRPPVPMPSHAAG 399
DB 360 GLPLGLH-SYNAGSINNLRVPAVAIPIPTPAAG 393

RESULT 5
O06341 PRELIMINARY; PRT; 393 AA.
AC O06341
DT 01-JUL-1997 (TREMREL. 04, Created)
DT 01-JUL-1997 (TREMREL. 04, Last sequence update)
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)
DE Hypothetical protein Rv3478.
GN Rv3478 OR MTCY13E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544 (1998).
DR EMBL; Z95390; CAB08702.1; -
DR PIR; C70568; C70568.
DR Tuberculist; Rv3478; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 52.6%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 7.8e-72;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 69 SSAGLWVAASPYVAVMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
DB 61 SSAGLWVAASPYVAVMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMT 120
QY 129 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPMTSAGG 188
DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATEALLPFEAPLITNPGG 180
QY 189 LLEQAAVEEASDTAAANOLMNNVPOALQOLAQPTGGTTPSSKLGGLWKTVPSPHSPLN 248
DB 181 LLEQAAVEEASDTAAANOLMNNVPOALQOLAQPTGGTTPSSKLGGLWKTVPSPHSPLN 240
QY 249 MYSMANNHMTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSS-----LGSSL 304
DB 241 VSSIANNHSMGTSGVSMNTLSSMLKGLAP-AAAQAVETAQAQNGVWAMSSLSGSLGSSL 299
QY 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMLG 364

```

```

Db 300 GSSGLGAGVAANLGRASVGLSVPPAWAANAQVTPAARALPLTSLTSAQAATPCHMLG 359
QY 365 GLPVGMQAGRAGGLGVLRLVPRPPVMPHSPAA 399
Db 360 GLPLGH-SVNAGSGINNLRVPAAYAIPTPTPAAG 393

RESULT 6
Q7TWFS
ID Q7TWFS PRELIMINARY; PRT; 393 AA.
AC Q7TWFS
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PE family protein.
GN PPE60 OR MB3505.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AP2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
RA "The complete genome sequence of Mycobacterium bovis."
RA Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL EMBL: BX248346; CAD95692.1; -.
RW Complete proteome.
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF42C0C95DAE7DD4 CRC64;

Query Match 52.6%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 7.8e-72;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 9 MVDFGALPPEINSARMYAGSGSLVAQAQWDSVASDLFSAASAFOSVWGLTVGSMIG 68
Db 1 MVDFGALPPEINSARMYAGSGSLVAQAQWDSVASDLFSAASAFOSVWGLTVGSMIG 60
QY 69 SSAGLVAQAASPVVWMSVTAGQAEITAAQVRVAAAYETAYCLTVPPVPIAENAEELMI 128
Db 61 SSAGLVAQAASPVVWMSVTAGQAEITAAQVRVAAAYETAYCLTVPPVPIAENAEELMI 120
QY 129 LIATNLLGQNTPAIAVNEAEYGEWIAQDAAMFGYAAATATATATALLPPEAPEMTSAGG 188
Db 121 LTATNLLGQNTPAIAVNEAEYGEWIAQDAAMFGYAAATATATATALLPPEAPEMTSAGG 180
QY 189 LLEQAAVEASDTAAANLMMNVPAALQALQAOPTGCTTPSSKLGGLMKTVPSPHSPISN 248
Db 181 LLEQAAVEASDTAAANLMMNVPAALQALQAOPTGCTTPSSKLGGLMKTVPSPHSPISN 240
QY 249 MVSMAHHMSMTNSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMS----LGSSL 304
Db 241 VSSIANNHMSMTNSGVSMNTLSSMLKGPAPAAAQAVTAAQNGVRAMS----LGSSL 299
QY 305 GSSGLGGVAANLGRASVGLSVPPAWAANAQVTPAARALPLTSLTSAAGRGQMLG 364
Db 300 GSSGLGAGVAANLGRASVGLSVPPAWAANAQVTPAARALPLTSLTSAQAATPCHMLG 359
QY 365 GLPVGMQAGRAGGLGVLRLVPRPPVMPHSPAA 399
Db 360 GLPLGH-SVNAGSGINNLRVPAAYAIPTPTPAAG 393

RESULT 7
O07175
ID O07175 PRELIMINARY; PRT; 355 AA.
AC O07175
DT 01-JUL-1997 (TrEMBLrel. 04, Created)

```

```

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Serine protease, putative).
GN PEPA OR RV0125 OR MTC1418B.07 OR MT0133.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H378V;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teklaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence."
RA Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.D., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains."
RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: Z96071; CAB09453.1; -.
DR EMBL: AE006925; AAK44357.1; -.
DR PIR: F70983; F70983.
DR TIGR: MT0133; -.
DR Tuberculist: Rv0125; -.
DR GO: GO:0004295; P:trypsin activity; IEA.
DR GO: GO:0007242; P:intracellular signaling cascade; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR008256; Peptidase_S1B_V8.
DR InterPro: IPR001940; Peptidase_S1C.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PRO0834; PROTEASES2C.
DR PRINTS: PRO0839; V8PROTEASE.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS00106; PDZ; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase, Hypothetical protein, Serine protease, Protease;
KW Complete proteome.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 32.9%; Score 989; DB 16; Length 355;
Best Local Similarity 98.0%; Pred. No. 3.9e-42;
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 397 AAGDIAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGVIDPN 456
Db 28 APAQAAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAGAGTGVIDPN 87
QY 457 GVVLTNHVTAGATDINAFSVGSGQYGVVDVGVYDRTQDVAVLQLRGAGLPSAALGGGV 516
Db 88 GVVLTNHVTAGATDINAFSVGSGQYGVVDVGVYDRTQDVAVLQLRGAGLPSAALGGGV 147
QY 517 AVGEPVAVANGSGGQGTTPRAVFCRVVVALGQTVQASDLSLTGAETLNLGTFQDAALQPGD 576

```

RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of *Mycobacterium bovis*";
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
DR ENBL; BX248340; CAD94520.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 39572 MW; 41944B6E547A8A0 CRC64;

Query Match 25.9%; Score 779.5; DB 16; Length 393;
Best Local Similarity 43.7%; Pred. No. 1.4e-31;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 10 VDFGALPPEINSARMYAGTGSASLVAAAQWDSVASDLFSAAGAFOSVWGLTVGSGWIGS 69
DB 1 MDFGALPPEVNSVRMYAGTGSAPMVAAAGSAWNGLAELSSAATGYETVITQLSSEGLWGP 60
QY 70 SAGLWVAASPYVAMSVMTAGQAEELTAAGVRVAAAAYETAYGLTVPPPVIAENRAELMIL 129
DB 61 ASAAEAIVAPYVAMNSAAAQAEQAATQARAAAFAEAFATVPPPLIANRASLMQL 120
QY 130 IATNLIGONTPAIVNAEYGENWADAAAMFYAAAATATATATATLLPFEAPEMTSAGGL 189
DB 121 ISTNVFGQNTSAIAAAEAQYGENWADSAAMYAYAGSSASASA-VTFPSTPPOIANPTAQ 179
QY 190 LEQAAAEVEASDTAAA-NQLNNVPOALQQLAQPTGGTTPSSKLGGLWK----- 237
DB 180 GTQAAAVATAGTAQSTLTEMITGLFVALQSTLPSLLQS-NGPLSWLWQILFGTTFNFT 238
QY 238 -----TVSPHSPISNNMVSMANNHMTNSGVSMNTLSMLKGFAPAAAAQAVQTAA 290
DB 239 SISALLTDLQPVASPFYNTGLPYFSGMGNFIOAAKTL-GLIGSAAPAVA-----AA 292
QY 291 QNGVRAKSSILGSSSGGGVAAALGPAASVGSLSVQAAWAA-ANQAVTPAARALPLT 349
DB 293 GDAAKGLGLGGMLG---GGPVAGLGNAAVSGKLSVPFVMSGPLPGSVTPGAAPLPVS 348
QY 350 SLTSAAREPGOMGLGPLVGOMGARAGGSLGVLRVPPRPYVMPHSPAAG 399
DB 349 TVSAAPEAAPGSLIGSLPL---AGAGGAGAGP-RVGFRPTVMAREPFGAG 393

RESULT 10
ID O53939 PRELIMINARY; PRT; 411 AA.
AC O53939;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1789 OR MT1838 OR MTV049.11.
OS *Mycobacterium tuberculosis*.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]_TaxID=1773;
SEQUENCE FROM N.A.
RP STRAIN=H37Rv;
RC MEDLINE=98295987; PubMed=9634230;
RA Gordon S.V., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Badcock K., Bisham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagele K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
RT complete genome sequence";
RT Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

```

RA Fleischmann R.D., Alland D., Eissen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.,
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL020201; CAA17711.1; ALT_INIT.
DR EMBL; AE007043; AAK46108.1; -.
DR PIR; G70929; G70929.
DR TIGR; MT1838; -.
DR TubercuList; RV1789; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 25.9%; Score 779.5; DB 16; Length 411;
Best Local Similarity 43.7%; Pred. No. 1.5e-31;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 10 VDFGALPPEINSAARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVMGLTVGSWIGS 69
DB 19 MDFGALPPEVNSVRMYAGPGSAPMVAAASAWNGLAELSSAATGYETVITQLSSEGWLGP 78
QY 70 SAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 129
DB 79 ASAMAEEAVPYVAMSAASAAQAEQATQARAAAAAFEAFAATVPPPLIAANRASLMOL 138
QY 130 IATNLGONTPAIVANEABYGEWMAODAAAFYAAATATATATATLPPFEAPEMITSAGGL 189
DB 139 ISTNVFGONTSAIAAEAYGEWMAODSAAMYAYAGSSASASA-VTPFGTTPQIANPTAQ 197
QY 190 LEQAAVEERASDTAAA--NOLNMVFPALQQLAQPTGGTTPSKLGLWK----- 237
DB 198 GTQAAAVATAAGTAQSTLTETMTGLPNALQSILTSPLLOSS-NGPLSLWQLGILFTPNFPT 256
QY 238 -----TVSPHSPISNMVSMANNMSTNSGVSMNTLSMLKGFAPAAAAAQVATAA 290
DB 257 SIGALLTDLQPYASFFNTBGLPYFYGIMGNFIQAKTL-GLIGSAAAPAAVA-----AA 310
QY 291 QNGVAMSSLSGSLGSGGVAANLGRAASVGSLSVPOANAA-ANQAVTPAARALPIT 349
DB 311 GDAAKGLPGLGMLG---GCPVAAGLGNAAASVGLSVPPVWSGSLPFGSVTPCAAPLPVS 366
QY 350 SLTSAERGGQMLGGLPVQCMGARAGGSLGVLRVPPRPYVMPHSPAAG 399
DB 367 TVSAAPAPAPGSLGGLPL-----AGAGGAGAGP-RYGFRTVVARPPFAG 411

RESULT 11
ID Q50320 PRELIMINARY; PRT; 361 AA.
AC Q50320;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 34kDa protein precursor.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J988/107;
RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
RT "Identification and characterisation of a putative serine protease
RT expressed in vivo by Mycobacterium avium subsp paratuberculosis."
Microbiology 140:1977-1982(1994).
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DB EMBL; Z23092; CAA80638.1; -.

PIR; S47170; S47170.
GO; GO:0008333; P:peptidase activity; IEA.
GO; GO:0004295; F:trypsin activity; IEA.
GO; GO:0007242; P:intracellular signaling cascade; IEA.
GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
InterPro; IPR009003; Cys_Ser_trypsin.
InterPro; IPR001478; PDZ.
InterPro; IPR001254; Peptidase_S1.
InterPro; IPR001940; Peptidase_S1C.
Pfam; PF00595; PDZ; 1.
Pfam; PF00089; trypsin; 1.
PRINTS; PR00834; PROTEASES2C.
SMART; SMO0228; PDZ; 1.
PROSITE; PS01016; PDZ; 1.
Hydrolase; Protease; Serine protease; Signal.
KW SIGNAL 1 38 POTENTIAL
SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match 25.9%; Score 778.5; DB 2; Length 361;
Best Local Similarity 66.0%; Pred. No. 1.4e-31;
Matches 155; Conservative 28; Mismatches 37; Indels 15; Gaps 2;

QY 362 MLGSLPVQCMGARAGGSLGVLRVPPRPYVMPHSPAAGDIAPALSDODREFADPALPLDP 421
DB 15 LVGLTVVGLGLGLGSGVG-----LAPASA--APSGLALDRPADRLAPIDP 59
QY 422 SAMVAQGPQVQVINTKLYNNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQ 481
DB 60 SAMVQGVQPVVINTKLYNNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQ 119
QY 482 TYGVVGVYDRTQDAVLQRLAGGLPSAAGGVAVGEVPMVAMSGSGGGTTPRAVPGR 541
DB 120 TYAVDVVGYDRTQDAVLQRLAGGLPSAAGGVAVGEVPMVAMSGSGGGTTPRAVPGR 179
QY 542 VVALGQTVQASDSLTGAETLNGLIQFDDAIQPCDGGPVVNGLGQVVGNTAAS 596
DB 180 VVALNQSVASITDITGAQENLGLIQADAPIKFGDSGPMVNSAGQVIGVDVTAAT 234

RESULT 12
QY99Q11 PRELIMINARY; PRT; 410 AA.
ID Q99Q11
AC Q99Q11 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE RV1808-like protein.
DE MYC1808 OR OV1808.
OS Mycobacterium microti.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1806;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MYC 94-2272, and OV254;
RA Sivadon V., Heym B., Mazancourt P., Gaillard J.-L.L.;
RT "PPE RV1808 orthologue of Mycobacterium microti."
Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF335180; AAK20894.1; -.
DR EMBL; AF335179; AAK20893.1; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
SQ SEQUENCE 410 AA; 40044 MW; 3B9DA3174655A5EA CRC64;

Query Match 25.2%; Score 756.5; DB 2; Length 410;
Best Local Similarity 42.4%; Pred. No. 2.1e-30;
Matches 182; Conservative 52; Mismatches 184; Indels 41; Gaps 11;

QY 9 MVDFGALPPEINSAARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVMGLTVGSWIG 68
DB 1 MLDGALPPEINSGRMVYAGPGSLAAALAAAWDAAALAAELYSAAASVGSSTIEGLTVAPMWG 60
QY 69 SSAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 128

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 9.70351 Seconds
(without alignments)
3198.204 Million cell updates/sec

Title: US-09-597-796c-10

Perfect score: 3007

Sequence: 1 HHMHHHHMDVFGALPPPIN.....SGGPPVNGLGQVGMNTRAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	1656.5	55.1	YD61 MYCTU	Q11031 mycobacteri
2	705	23.4	Y892 MYCTU	Q10813 mycobacteri
3	688.5	22.9	Y102 MYCTU	O53951 mycobacteri
4	448	14.9	YF48 MYCTU	Q10778 mycobacteri
5	445	14.8	Y442 MYCTU	P42611 mycobacteri
6	426.5	14.2	Y478 MYCTU	Q10540 mycobacteri
7	365	12.1	SRA MYCTU	Q07297 mycobacteri
8	338	11.2	YU18 MYCTU	P31500 mycobacteri
9	334	11.1	Y096 MYCTU	Q10892 mycobacteri
10	333.5	11.1	YU21 MYCTU	O53268 mycobacteri
11	262.5	8.7	YVTA BACSU	Q95911 bacillus su
12	233.5	7.8	YX29 MYCTU	O06246 mycobacteri
13	225.5	7.5	YX29 MYCTU	Q91010 arabidopsis
14	225	7.5	DEGS_ECOLI	P31137 escherichia
15	225	7.5	DEGF_CHLEN	Q92670 chlamydia p
16	224.5	7.5	DEGI_ARATH	O43464 homo sapien
17	222	7.4	DEGI_ARATH	O22609 arabidopsis
18	222	7.4	DEGI_ECOLI	P39099 escherichia
19	221.5	7.4	DEGS_MOUSE	O94195 mus musculu
20	221	7.3	HRA2 BACSU	O34358 bacillus su
21	221	7.3	DEGF_BRUSU	Q44597 brucella su
22	219.5	7.3	YXKA BACSU	P39668 bacillus su
23	218.5	7.3	DEGP_BUCAI	P57322 buchiera ap
24	217.5	7.2	YX25 MYCTU	O50703 mycobacteri
25	217	7.2	HTOA_HAEIN	P45129 haemophilus
26	215	7.1	HTRA_LACHE	O82477 lactobacill
27	215	7.1	DEGP_EICGN	Q92341 rickettsia
28	213	7.1	DEGF_BRUNE	Q89532 brucella me
29	211	7.0	DEGP_SALTY	P26982 salmonella
30	208.5	6.9	DEGP_ECOLI	P09376 escherichia
31	208.5	6.9	DEGP_CHLTR	P18584 chlamydia t
32	207.5	6.9	DEGP_BUCAP	O85291 buchiera ap
33	206.5	6.9	DEGP_CHLMU	Q9p197 chlamydia m

ALIGNMENTS

RESULT 1

ID	YD61_MYCTU	STANDARD;	PRT;	396 AA.
AC	Q11031;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical PPE-family protein Rv1361c/MT1406.			
GN	Rv1361c OR MT1406 OR MTCY02B10.25C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RA	Medline=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Hornesby R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornesby T., Jagels K., Krogh A., McLean J., Moulé S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Medline=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	J. Bacteriol. 184:5479-5490(2002).			
CC	-1- SIMILARITY: Belongs to the mycobacterial PPE family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z75555; CAA99966.1; -			
DR	EMBL; AE007013; AAK45669.1; -			
DR	PIR; H70741; H70741.			
DR	TIGR; MT1406; -			
DR	TubercuList; Rv1361c; -			
DR	InterPro; IPR000030; Microbac_PPE.			

Q52894 rhizobium m
O05942 rickettsia
P54925 bartonella
Q93372 rattus norv
P83105 homo sapien
P15502 homo sapien
Q91a06 lactococcus
Q92743 homo sapien
P04985 bos taurus
Q91118 mus musculu
P83110 homo sapien
P44947 haemophilus

DR Pfam: PF00823; PPE: 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 158 159 TA -> AT (IN REF. 2).
 SQ SEQUENCE 396 AA; 40015 MW; 6AFAD0D7B5F668D0 CRC64;

Query Match 55.1%; Score 1656.5; DB 1; Length 396;
 Best Local Similarity 85.1%; Pred. No. 5.5e-77;
 Matches 338; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 9 MVDFGALPPPEINARMYAGPGSASLVAAAQWDSVADLFSASATQSVVWGLTVGSWIG 68
 DB 1 MVDFGALPPPEINARMYAGPGSASLVAAAQWDSVADLFSASATQSVVWGLTVGSWIG 60

QY 69 SSAGLVAAASPVVWMSVTTAGQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128
 DB 61 SSAGLVAAASPVVWMSVTTAGQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120

QY 129 LIATNLGQNTPAIAVNEAEYGENWQADAAAFYGAATATATATATLLPPEAPEMTSAGG 188
 DB 121 LIATNLGQNTPAIAVNEAEYGENWQADAAAFYGAATATATATALLPPEAPEMTSAGG 180

QY 189 LLEQAAVEASDTAAANQIWNVPAQLQALQAOPTGGTTPSSKGLGLWKTVPVPHRSPIN 248
 DB 181 LLEQAAVEASDTAAANQIWNVPAQLQALQAOPTGGTTPSSKGLGLWKTVPVPHRSPIN 240

QY 249 MVSMNNHMTNSGVSMNTLTSLMLKGFAPAAAQAVOTAAQNGVRVMS-----LGSSL 304
 DB 241 IVSMNNHMTNSGVSMNTLTSLMLKGFAP-AAAQAVETAAQNGVQVMSLGLSSQLGSSL 299

QY 305 GSSGLGGVAAANLGRAASVCSLVQPAWAAANQAVTPAARALPLTSLTSAEREGPQMLG 364
 DB 300 GSSGLGGVAAANLGRAASVCSLVQPAWAAANQAVTPAARALPLTSLTSAEREGPQMLG 359

QY 365 GLPVGQMGARAG--GGLSGLVLRVPPRPVPMHPSPAAG 399
 DB 360 GLPVGQMGARAG--GGLSGLVLRVPPRPVPMHPSPAAG 396

RESULT 2
 YS92 MYCTU STANDARD; PRT; 408 AA.
 AC Q10613;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV2892C/MT2959/MB2916c.
 GN RV2892C OR MT2959 OR MTCY274.23C OR MB2916C.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 ON NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duchoy S., Gordin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -I- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 274024; CA98377.1; -;
 DR EMBL; AE007119; AAK47285.1; -;
 DR EMBL; BX248344; CAD96603.1; -;
 DR PIR; G70925; G70925.
 DR TIGR; MT2959; -;
 DR TubercuList; RV2892C; -;
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 56 POTENTIAL.
 SQ SEQUENCE 408 AA; 41469 MW; 3E3D1F20D7827199 CRC64;

Query Match 23.4%; Score 705; DB 1; Length 408;
 Best Local Similarity 41.8%; Pred. No. 5.2e-28;
 Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;

QY 10 VDFGALPPEINARMYAGPGSASLVAAAQWDSVADLFSASATQSVVWGLTVGSWIGS 69
 DB 1 MDFGVLPEINSGRMVYAGPGSGPWWMAAAAADSLAELGLAAGCYRLAISELTGAYWAGP 60

QY 70 SAGLVAAASPVVWMSVTTAGQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 129
 DB 61 AAASWVAATVPYVAVLGSATAGQAFQAGQWQRAAAAAYELAFAMTVPPVVAANRALVAL 120

QY 130 IATNLGQNTPAIAVNEAEYGENWQADAAAFYGAATATATATLLPPEAPEMTSAGGL 189
 DB 121 VATNFFGQNTPAIAATEAQIAEMWQAQDAANAYAGSAATAT-ELTPFTAAPVTTSPAAL 179

QY 190 LEQAAA-----VEEASDTAAANQLMNN-----VP--QALQO--LAQPTQGTTPSSKL--- 232
 DB 180 AQAAAATVSTVTPPLATTAAVAPQLLOQLSSTLSIPWYSAQQWLAENLLGTPPNRTIV 239

QY 233 -----GGLWKTVPVPHRSPISNMVSMNNHMTNSGVSMNTLTSLMLKGFAPAAAA 283
 DB 240 RLLGLSYFDEGL-----LQFEASLAQQA-FGTPGGAG--DSGSSVLDSKGTIFA 287

QY 284 QAVQTAAQNGVRMSSL--GSSLGS-----SGLGGVAAANLGRAASVCSLS 327
 DB 288 -----GPRASPSVAGGAVGVQTPQPYWYMWALDRESIGGSVAALGKGSAGLS 338

QY 328 VPQAWAAANQAVTPAARALP---LTSLSAEREGPQMLGGLPVQMGARAGGLSGVLR 384
 DB 339 VPFDWAAARWANPAMWLPDGDVTLRGTAENA---LLRGFFMASAGQSTGGGF--VHK 393

QY 385 VPPRPVPMHPSPAAG 399
 DB 394 YGFRILAVMQRPFAG 408

RESULT 3
 ID Y102 MYCTU STANDARD; PRT; 463 AA.
 AC OS3951;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv1802/MT1851/MD1830.
 GN Rv1802 OR MT1851 OR MV049.24 OR MS1830.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RL laboratory strains.";
 RN J. Bacteriol. 184:5479-5490 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 DR EMBL; AL022021; CAAL17723.1; -
 DR EMBL; AE007044; AAK46123.1; -
 DR EMBL; BX248340; CAD94533.1; -
 DR PIR; C70931; C70931.
 DR TIGR; MT1851; -
 DR TubercuList; Rv1802; -
 DR InterPro; IPR00030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 401 401 S -> L (IN REF. 2).

SQ SEQUENCE 463 AA; 46021 MW; EB64B28BF09FA551 CRC64;
 Query Match 22.9%; Score 688.5; DB 1; Length 463;
 Best Local Similarity 42.8%; Pred. No. 4e-28;
 Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;
 QY 10 VDFGALPPEINSARMYAGPGSASLVAAQOMWDSVASDLFSAAAFQSVVWGLTVGSWIGS 69
 DB 1 MDFGVLPEINSGRMYAGPGSGPMLAAAAWDGLATELQSTAADYGSVISVLT-GVWSGQ 59
 QY 70 SAGLMVAASPYVAMSVTGAQELTAARVRAAAAYETAYGLTVPPPIAENRAELMIL 129
 DB 60 SSGTMAAAAPYVAMMSATAALAREFAAQAAGAAAAYEAFATVPPPVAAANRAELAVL 119
 QY 130 IATNLLGONTTALAIVNEAEYGEWMAQDAAMAFGAAATATATATLLPPEEAPMTSAGCL 189
 DB 120 AATNIFGQNTGATIAAAEAARYAEMWQDAAMAYGAGSSSVAT-QVTTPAAPPTTNAAGL 178
 QY 190 LEQAAVEEASDTAAANQNMNVFQALQLOAQPTGGTTPPSKLGGLMKTVS--PHRSPI- 246
 DB 179 ATQGVAVQAAGCAGAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLMMNAVTVGVYASSVY 235
 QY 247 -----SNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAQVTAAGNGVRA 296
 DB 236 NSMLGLGFAESKRVLPANDTVISIFGMVQFKFNPVTPNPDLIPK----- 283
 QY 297 MSSLGSSLG-----SSGLGG---GVANLGRAASVGSLSVFPQAAANQAVTPAARALPL 348
 DB 284 -SALGAGLGLRSATSSGLGSTAPASAGASQAGSVGMSVFPSSAAATPAIRTVAAVFSS 342
 QY 349 TSLTS--AERGPQQL-----GGLPVQMGARAGGGLSGVLRV 385
 DB 343 TGLQAVPAAAISEGLSLQMALASVAGGALGAAARATGGLGGRV 389
 RESULT 4
 YF48 MYCTU
 ID YF48 MYCTU STANDARD; PRT; 678 AA.
 AC Q10778;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv1548c/MT1599.
 GN Rv1548C OR MT1599 OR MCV48.17.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RL laboratory strains.";
 RN J. Bacteriol. 184:5479-5490 (2002).

Best Local Similarity 31.9%; Pred. No. 8e-16; Matches 135; Conservative 49; Mismatches 181; Indels 58; Gaps 12;	
QY 12	FGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIGSSA 71
Db	6 FAWLPPEINSAALFAGPGSGLPAAATANGELAEELIASLGSVTSLSLTGAWLGPSSA 65
QY 72	GLMVAASPYVAMSVYTAGQAEITAAQVRVAAAYETAYGLTVPPVIAENRAELMILIA 131
Db	66 AAMVAVATQYLAWLSTAAQAQAEQAQAAMAIATAPFAALAAATVQPAVVAANRGLMQLLAA 125
QY 132	TNLLGONTALAVNEAEYGEWAODAAAMFGVAAATATATATATALLPFEAPEMTSAG---- 187
Db	126 TNWFGQNALMDVDEAYFQWALDVAANWAGYHFDASAAVAQAPQOQV--LRNLGIDIG 183
QY 188	--GLLEQAAAEVBEASDTAAANLMNVPOALQAOPTGCTPSSKLG-----GLW 236
Db	184 KXGQINLFGTSGNIGNNNIGNNNIGNNNIGNNNIGNNNIGNNNIGNNNIGNNNIGNNN 243
QY 237	KIVS-----PHRSPISNMVSMANNEMMTNSGVMTLSSMLKGFAPAAQAQAVQ 287
Db	244 NTGSGNIGFITGDHOMGFGGFGSGGN-IGFGNSGTGNVGLFNS----- 287
QY 288	TAAQNGVRAMSSILGSSGLSGGGVAAANLGRAASVGSLSVQAAVAAANQAVTPAARALP 347
Db	288 GSGNIGTSGNIGSNGISGTGT---INAGLG---SAGSLNT-SFNWAGNNAALGSAAGS 340
QY 348	LTSLSAAERGGQOM-----LGLPVGQMGARAG--GGLSGVLR--VPRPYVMHPSPAA 398
Db	341 EAAVSSAGYATGCMSTAALSSGILASALGSGTGLQHLGIANVLNGLTNPVPAAPASAPV 400
QY 399	GDI 401
Db	401 GGL 403
RESULT 6	
Y878 MYCTU	STANDARD; PRT; 443 AA.
ID Y878 MYCTU	
AC Q10540;	
DT 01-OCT-1996 (Rel. 34, Created)	
DT 01-OCT-1996 (Rel. 34, Last sequence update)	
DT 10-OCT-2003 (Rel. 42, Last annotation update)	
DE Hypothetical pPE-family protein Rv0878c/MT0901.	
GN Rv0878C OR MT0901 OR MTCY31.06C.	
OS Mycobacterium tuberculosis.	
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX NCBI_TaxID=1773;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=H37Rv;	
RC MEDLINE=98295987; PubMed=9634230;	
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,	
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,	
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,	
RA Oliver S., Osborne J., Quail W.A., Rajandream M.A., Rogers J.,	
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,	
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;	
RT "Deciphering the biology of Mycobacterium tuberculosis from the	
RL complete genome sequence."	
RL Nature 393:537-544(1998).	
RN [2]	
RP SEQUENCE FROM N.A.	
RC STRAIN=CDC 1551 / Oshkosh;	
RX MEDLINE=22206494; PubMed=12218036;	
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,	
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,	
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,	
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,	
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;	

"Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";	
RL J. Bacteriol. 184:5479-5490(2002).	
CC SUBCELLULAR LOCATION: Integral membrane protein (potential).	
CC SIMILARITY: Belongs to the mycobacterial PPE family.	
CC	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC or send an email to license@sib-sib.ch).	
CC	
DR EMBL; Z73101; CA97385.1; --	
DR EMBL; AE006977; AAK45143.1; ALT_INIT.	
DR PIR; C70780; C70780.	
DR TIGR; MT0901; --	
DR TubercuList; Rv0878c; --	
DR InterPro; IPR000030; Microbac_PPE.	
DR InterPro; IPR002989; Mycobac_Pentapep.	
DR Pfam; PF01469; Pentapeptide_2; 4.	
DR Pfam; PF00823; PPE; 1.	
DR Hypothetical protein; Transmembrane; Repeat; Complete proteome.	
FT TRANSMEM 15 35 POTENTIAL.	
FT TRANSMEM 38 58 POTENTIAL.	
FT TRANSMEM 59 79 POTENTIAL.	
FT TRANSMEM 181 201 POTENTIAL.	
FT DOMAIN 64 73 POLY-ALA.	
FT DOMAIN 81 115 ALA-RICH.	
FT DOMAIN 231 270 4 X 10 AA APPROXIMATE REPEATS.	
FT SEQUENCE 443 AA; 43592 MW; C58BEC607F0675E2 CRC64;	
Query Match 14.2%; Score 426.5; DB 1; Length 443;	
Best Local Similarity 31.2%; Pred. No. 6.2e-15;	
Matches 125; Conservative 56; Mismatches 132; Indels 87; Gaps 14;	
QY 10	VDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVG----SW 66
Db	1 MNFMVLPPEVNSARIYAGAPAPMLAAAVADWGLAELGMAAAAFSLISGLTAGPGSAW 60
QY 67	IGSSAGLWMAAASPYVAMSVYTAGQAEITAAQVRVAAAYETAYGLTVPPVIAENRAEL 126
Db	61 QGPAAAAAANAAPYLSWLNAAATARAAGAAAGAKAAAVYEAARAATAHALVAANRQL 120
QY 127	MILIATNLLGONTALAVNEAEYGEWAODAAAMFGVAAATATATATATALLPFEAPEMTSA 186
Db	121 LSLVLSNLFQNLPAIATAATEASYEQLWAQDVAAMVYHGGASTVASOLTPWQQ----- 173
QY 187	GGLEQAAAEVBEASDTAAANLMNVPOALQAOPTGCTPSSKLGGLMKTVPSPHSPI 246
Db	174 --LLSVLPVVTAPAGAV-----GVPA--JAI PALGV---ENIG-----V 208
QY 247	SNMV---SMANNHMTNSG-----VSMTN-----TLSSMLKGFAPAAAAQAV 286
Db	209 GNFLGIGNIGNNVGSGTGDYFNFGIGNGNANLGNIGNGNANLGNIGNGNAGFFNFGNGNDG 268
QY 287	QTAQNGVRAMSSILGS--SLGSSGLGGVAA--NLGRAASVGSLSVQAAVAAANQAVTAA 343
Db	269 NTNFGSGNAGFLNIGSGNEGSLNFGNAGDDNTG-----WNSGD----- 309
QY 344	RALPLATSLTSAERPGCQMLGGL--PYGQMGARAGGGLSG 381
Db	310 -----TWTGGFNSGDLNTGIGSPVTQGVANSFGNTG 341
RESULT 7	
SRA_MYCLE	STANDARD; PRT; 408 AA.
ID SRA_MYCLE	
AC Q07297;	
DT 01-NOV-1995 (Rel. 32, Created)	
DT 01-NOV-1995 (Rel. 32, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	

DE Serine-rich antigen (25kL) (45 kDa protein).
GN SPA OR ML0411 OR MLC1383.14.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93293328; PubMed=8478104;
RA Vega-Lopez F., Brooks L.A., Dockrell H.M., de Smet K.A.L.,
RA Thompson J.K., Hussain R., Stoker N.G.;
RT "Sequence and immunological characterization of a serine-rich antigen
RT from Mycobacterium leprae";
RL Infect. Immun. 61:2145-2153(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95020554; PubMed=7934845;
RA Rinke de Wit T.F., Clark-Curtiss J.E., Abebe F., Kolkon A.H.J.,
RA Jonson A.A.M., Thole J.E.R.;
RT "A Mycobacterium leprae-specific gene encoding an immunologically
RT recognized 45 kDa protein";
RL Mol. Microbiol. 10:829-838(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=TN;
RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby I., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus";
RL Nature 408:1007-1011(2001).
RL Nature 408:1007-1011(2001).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL; U00015; AAC43220.1; --
CC EMBL; X68431; CAA48480.1; --
CC EMBL; Z21952; CAA79950.1; --
CC EMBL; Z37179; CAB09338.1; --
CC EMBL; AL583918; CAC229919.1; --
CC PIR; C86960; C86960.
CC PIR; S33522; S33522.
CC PIR; S39872; S39872.
CC Leprena; ML0411; --
CC InterPro; IPR000030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1
CC Antigen; Repeat; Complete proteome.
KW Antigen; Repeat; Complete proteome.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 209 235 2 X 6 AA REPEATS OF S-V-A-Q-S-E.
FT REPEAT 209 214 1.
FT REPEAT 230 235 2.
FT REPEAT 132 132 T -> S (IN REF. 2).
FT CONFLICT 189 189 S -> L (IN REF. 2).
FT CONFLICT 191 191 H -> D (IN REF. 2).
FT CONFLICT 292 292 P -> L (IN REF. 2).
SQ SEQUENCE 408 AA; 42466 MW; 5C0C2B80D6E6A9D8 CRC64;

Query Match 12.1%; Score 365; DB 1; Length 408;
Best Local Similarity 26.2%; Pred. No. 7.2e-12;
Matches 117; Conservative 74; Mismatches 168; Indels 88; Gaps 12;

QY 9 MYDFGALPPEINSARMYAGPGSASLVAAACQWDSVADSLFSAASAFQSVVWGLTVGSWIG 68
DB 1 MFDEFWVSPVNAFLMGRGPGSTPLWGAABAWISLAEQLMEAAQEVSDTTIVVAVPASFAG 60
QY 69 SSAGLMVAASPYVAVWSVTAGQALTAACVRAAAAYETAYGLTPPVPIAENRAELMI 128
DB 61 ETSMDLASRVSTTFVWLDGNAENGLIARVHVAVAFEAERAGWVPLLTVLGNIIHTMA 120
QY 129 LIATNLLQNTPAIYVNEAYGEMWQDAAMFYAAATATATATATLTPFEAEPMTSAGG 188
DB 121 LKAINWFGVSTTVAALEADYDLMMVQNSTAMTYRTVLTRETCKMENFEPAPQLVSR-Y 179
QY 189 LLEQAAVAEASDTAAANQLMNNVPOALQLOAQ----- 221
DB 180 CMDRRDSVNSFHSSSDSLYESIDNLYDSVAOSEEHGSDMSQSYNTCCSVQASELCDS 239
QY 222 ----PTO-----GTPSSKLGGLMKTVPSPHSRPSINMVMANNHMTNSGVSMNT 269
DB 240 PFGTSSQSSQNDLSATSLTQQLGGL-----DSIISSASASLLATNS--ISSST 286
QY 270 LSSMLKGFAPAAQAQVQTAQNGVRAMSSGLSGSLGGGGVAAN-----LGRAASVG 324
DB 287 ASSIM----PIVASQVTTLTGRSQV-AVERKMIOSISSTVSVDDVAASKVAGVQAVSVG 341
QY 325 SLSVPOAWAANAQVTPAARALP-LTSLTSAABERGPGQMLGGLPVQGMGARAGGGISGV 382
DB 342 ALRVPEWNTASQPMATASVPAACSAITVA-----VSGPLEGV 381
QY 383 LRVPFRPYVMPHSPAAGD-IAPPALSQ 408
DB 382 TQ--PAEVLTSVAGSGGTGPPAFNE 406

RESULT 8
YU18 MYCTU STANDARD; PRT; 434 AA.
AC P31500; OS3265;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3018C/MT3098/MT3101.
GN RV3018C OR MT3098/MT3101 OR MTV012.32C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Biglmeier K., Gas S., Chillingworth T., Connor R.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Gentles S., Hamlin N., Holroyd S.,
RA Davies R., Devlin K., Feltwell T., Mearns J., Moule S., Murphy L.,
RA Hornsby T., Jagels K., Krohn A., Maclean J., Rajandream M.A., Rogers J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Squares R.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490(2002).

```

[3]
RN   SEQUENCE OF 160-374 FROM N.A.
RC   STRAIN=Isolate 50410;
RA   Pakki A.H., Dale J.W.;
RL   Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC   -1- SIMILARITY: Belongs to the mycobacterial ppe family.
CC   -1- CAUTION: In strain Oshkosh the gene for this protein is
CC   interrupted in position 307 by an IS6110 element.
CC   -1- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
CC   reductase.
CC   -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC   in positions 294; 337 and 355.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; AL021287; CAA16103.1; -
DR   EMBL; AE007129; AAK47427.1; ALT_SEQ.
DR   EMBL; AE007129; AAK47430.1; ALT_SEQ.
DR   EMBL; X59271; CAA4196.1; ALT_FRAME.
DR   F08; E70857; E70857.
DR   TIGR; MT3098; -.
DR   TIGR; MT3101; -.
DR   TubeXculist; RV3018c; -.
DR   InterPro; IPR000030; Microbac_ppe.
DR   Pfam; PF00823; PPE; 1.
DR   Hypochemical protein; Complete proteome.
KW   HYPOCHEMICAL PROTEIN; COMPLETE PROTEOME.
SQ   SEQUENCE 434 AA; 43029 MW; 41D6734BD389DD6 CRC64;

Query Match          11.2%; Score 338; DB 1; Length 434;
Best Local Similarity 27.2%; Pred. No. 1.8e-10;
Matches 123; Conservative 54; Mismatches 211; Indels 64; Gaps 12;

QY   14 ALPPEINSRMYAGPGSASIVAAQAQWDSVASDLFSAASAFQSVVGLTVGSGWIGSSAGL 73
DB   8 ASPEVHSALLSKGPGFGLCAAGHSALSYAAVAQELSVVAAGVGWQGSDEL 67

QY   74 MYAASPYVAMSVTGAQBELTAQVVRVAAAAYETAYGLTVPPVPAENRAELMILLIATN 133
DB   68 FVAAYYPVYAWLVQASDASAAAGEHAAAGYVVCALAEMPTLPELAANHLTHAVLVATN 127

QY   134 LLGONTPTAIIVNAEYGEWMAQDAAMAFGVAATATATATLLFFEPAPENTSAGGLEQA 193
DB   128 FFGINTIPTALNADYVRVMVQAATVMSAYEVVUGAALVATPTGTGAPVIVKEP----- 181

QY   194 AAVEASDTAAAN-----QLMNNVPQALQLAQPTQGTTPSSKLGGLWKTVPSPH 242
DB   182 --ANEASNAVAATITPFPWHEITVQFLEETFAAYDQVLSALLSELPA--VAWVWFQLFVD 237

QY   243 ---RSPTSNMVMNNHMSMTNSGVSMNTLLSMLKGFAPAAAQAQVOTAAQNGVAMSS 299
DB   238 ILGFIILGFIITILASNAQLLTFEAINASYAVAGLLYAIA-GVTDIVVEWIGMLFGVPL 296

QY   300 LGSSL-----GSSGLGG--GVAA--NLGRAASVGSLSVPQAWAANQAVTPAARA 345
DB   297 LGGPELLGALAAAIVVPGVAGIAGVAGLAALPAVGAAAGAPAAALYGVSVAPSGGVSPQAR- 355

QY   346 LPLTSLTSAERGGQMLGGLPVGQMGARAGG--GLSGVLRVPPRVVPMHPHSPAAGDIAP 403
DB   356 ----LVSAYEPAPASTSVSVLASDRGAGLGFVGTAG-----KESVQGPAG 397

QY   404 PALSQDRFADFPAFLDPSAMVAQVGFVYNI 435
DB   398 LTVLADFEFGDGAQVPVPLPGSW----GPDLVGV 425

```

RESULT 9
Y096 MYCTU

[illegible]

Q7 14 ALPPEINARMYACPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWITGSSAGL 73

Db 2 AIPPEVHSGLLSAGCGPGSLLVAAQQWQLSDQYALACAEGLGQLLGEVQVSSWQGTAAQ 61
 QY 74 MVAAGPYVAVMVGSTAGQAEITAAQVRAAAAYETAYGLTVPVPIAENRAELMILIAIN 133
 Db 62 YVAAHGYLAWLEQTAINSVTAQVRAAAAYCSALAAAMPTPAELAAHAIHGVLIATN 121
 QY 134 LLGQNTFAIVNEAEYGEVMAQDAAMFGVAAATATATATATLTPFEAPMTSAGGLLEQA 193
 Db 122 FFGINTVPIALNEADYVRVWLQADTMAAYQVADAATVAVPSTQAPPPIRPGG----- 176
 QY 194 AAVEASDT-----AAANQNMNVPOALQOLAQTO-----CTTPSSKL----- 232
 Db 177 -----DAADTRDLVLSIGQIRI-----LDFIANPYKVFLEFFEQFGFSAVTVLALVAL 229
 QY 233 ---GGLW-----KTVSPHRSPISNMVSMANNHMTNMTNLTSLMLK-----GF 277
 Db 230 QLYDFLWYPYVAGVGLLLPFPFP-----TSLALTALSALIHLLNPPAGL 275
 QY 278 AAAAAQVTAQVTAQNGVRAMSSLSGSLGGVAAVNLGRAASVGSLSVPCAAANQ 337
 Db 276 LPTAAA-----LPGDQMGANLA-----V 294
 QY 338 AVTPAARALPLTSLTSAAGPGQMLGGLFVQMGARAGGSLGVLRVPPRYVMPHSPA 397
 Db 295 AVTPATAVP-----GGSPPTSNPAPAAPSSNSVGSASAAFGI---SVA 335
 QY 398 AGDIAPPALSQDRPA---DFPALDPSAMVAQVGPQVNVNKLGYNNVAGAGTGVIDP 455
 Db 336 VPGIAPPVGSVSGPKAGTKSPDTAATLATAGAAARPGIARHRRKRSSEVGG-----IRGY 390
 QY 456 NGVVLTNHVIAGATDINAPSVSGGTYGVVGYDRTQDVALQLRGAGGLPSAAIGGG 515
 Db 391 RDEFLDATAVDAATVPAPANNAG---SQAGTILGF-----AGTAPTIS---G 433
 QY 516 VAVG 519
 Db 434 AAAG 437

RESULT 10
 YU21_MYCTU STANDARD; PRT; 435 AA.
 AC O53268; O53269;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV3022c/RV3022c/MT3106.
 GN RV3022c/RV3022c OR MT3106 OR MT3106.35C/MT3106.36C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultston J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Deicher A., Ufferback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL J. Bacteriol. 184:5479-5490 (2002).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
 frameshift in position 82.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL021287; CAAL16106.1; ALT_FRAME.
 CC EMBL; AL021287; CAAL16107.1; ALT_FRAME.
 CC EMBL; AE007129; AAK47435.1; -
 CC TIGR; MT3106; -
 CC Tuberculan; RV3021c; -
 CC Tuberculan; RV3022c; -
 CC InterPro; IPR000030; Microbac_PPE.
 CC Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 299 299 G -> A (IN REF. 2).
 FT CONFLICT 317 320 LAGV -> VTGL (IN REF. 2).
 FT CONFLICT 326 326 L -> V (IN REF. 2).
 SQ SEQUENCE 435 AA; 42876 MW; 3B157643EAA8484A CRC64;
 Query Match 11.1%; Score 333.5; DB 1; Length 435;
 Best Local Similarity 25.3%; Pred. No. 3e-10;
 Matches 123; Conservative 59; Mismatches 172; Indels 133; Gaps 14;
 QY 14 ALPPEINARMYAGPGSASLVAAQWMDVSADLSAASAFQSVWGLTVGWSIGSAGL 73
 Db 8 ASPPEVHSHALLSAGPGPSLQAAAAGWSALSAAVAAQSLVVVAAVAGVWQGSSEL 67
 QY 74 MVAAGPYVAVMVGSTAGQAEITAAQVRAAAAYETAYGLTVPVPIAENRAELMILIAIN 133
 Db 68 FVAAVYFVAVLWVQASADSAAGAAAGVYCALAEMPTLPFLAANLTHAVLATN 127
 QY 134 LLGQNTFAIVNEAEYGEVMAQDAAMFGVAA----- 165
 Db 128 FFGINTVPIALNEADYVRVWVQAATVMSAYEAVVGAALVATPHTGPAVIVKPGANEASN 187
 QY 166 ATATATATLTPFE-----APEMTSAGLLEQAAVEASDSTAANOLMNVPOALQ 217
 Db 188 AVAAATITPFPFGLAKFLEMAQAFTVEGELINKSAEAWAVGFVELITGLVNEP----- 243
 QY 218 QLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSM-----ANNH 256
 Db 244 -----NLV-----LTCMDMFPATVGFALGVFLVPLLEFAVLE 278
 QY 257 MSMTNSGVSMNTLSS-----MLKGFAPAAAQAQVQTAONGVRAMSSLSGSLSSGL 309
 Db 279 LAILSIGWISNIFGAIPVLGGLLALAAVVGAGLAGVAGLAALPAVGAAGAP-- 336
 QY 310 GGGVAAVNLGRAASV-GSLSVFQAAVAAVQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368
 Db 337 -----AALVGSVAVPVGSGVSPQALVS--AVEPAPASTSVSL--ASDRGAGAL--GF-V 385
 QY 369 QMGARAGGSLGVLRVPPRYVMPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAQV 428
 Db 386 GTAGKESVGPAGL-----TVLADEFQDGPVPMPLPGSM----- 419
 QY 429 GPQVNI 435
 Db 420 GPDLVG 426


```

RESULT 11
ID YVTA_BACSU STANDARD; PRT; 458 AA.
AC Q9R9T1; Q35021; Q35039;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease yvta (EC 3.4.21.-).
GN YVTA OR YVTB OR BSU33000.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=20158875; PubMed=10692364;
RA Noone D., Howell A., Devine K.M.;
RT "Expression of ykda, encoding a Bacillus subtilis homologue of HtrA,
RT is heat shock inducible and negatively autoregulated.";
RL J. Bacteriol. 182:1592-1599 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98015415; PubMed=933931;
RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
RT "Sequencing of regions downstream of addA (98 degrees) and citG (289
RT degrees) in Bacillus subtilis.";
RL Microbiology 143:3305-3308 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=980404033; PubMed=9384377;
RA Kunst F., Gysasawa N., Mosser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Borriero S., Boursier L., Brans A., Braun M., Brignell N.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Frits C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guileppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Tachuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
RN [4]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20576168; PubMed=11133960;
RA Noone D., Howell A., Collier J., Devine K.M.;
RT "ykda and yvta, HtrA-like serine proteases in Bacillus subtilis,
RT engage in negative autoregulation and reciprocal cross-regulation of
RT ykda and yvta gene expression.";
RL J. Bacteriol. 183:654-663 (2001).
CC -!- FUNCTION: May be involved in processing, maturation, or secretion
CC of extracellular enzymes.

```

```

CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- INDUCTION: Induced by heat shock during exponential growth and by
CC heterologous amylases at the transition phase of the growth cycle.
CC Negatively regulates its own expression.
CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
CC of htrA, especially during stress conditions.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 P02/DHR domain.
CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to
CC frameshifts in positions 87 and 246 that produce two separate
CC ORFs.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF188296; AAF03153.1; -
CC EMBL; Z93941; CAB07968.1; ALT_FRAME.
CC EMBL; Z93941; CAB07969.1; ALT_FRAME.
CC EMBL; Z99120; CAB15290.1; ALT_FRAME.
CC Subtilist; BG14155; yvta.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_SIC.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF00083; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS50106; PDZ; 1.
CC Hydrolase; Protease; Serine protease; Heat shock; Transmembrane;
CC Complete proteome.
CC DOMAIN 1 71 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 72 92 POTENTIAL.
CC DOMAIN 93 458 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 356 440 PDZ.
CC ACT_SITE 187 187 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 217 217 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 298 298 CHARGE RELAY SYSTEM (POTENTIAL).
CC SEQUENCE 458 AA; 48717 MW; 77551045A865ASCD CRC64;
CC -----
CC Query Match 8.7%; Score 262.5; DB 1; Length 458;
CC Best Local Similarity 29.1%; Pred. No. 1.2e-06;
CC Matches 85; Conservative 52; Mismatches 116; Indels 39; Gaps 14;
CC -----
CC QY 326 LSVFQAAANQAVTPAARALPLTSLTSAERGGQMLGGI-PVGQMGARAGGSL-GVL 383
CC Db 38 LDAPVTEAGRQ---ETASALEMEKQETAVKKKRRRAWLSPI--LGGIIGGLMLGI- 91
CC QY 384 RVPRPYVMPHSPAAAGDIAPALSDQRFADFPALPLDPSA---MVAQVGPVVNI-NTK 438
CC Db 92 ----APVLPDSDQATETA-SANKQVQSDNFTTAPITNASNIADMDVDELTPTVIGISNIQ 146
CC QY 439 LGYNAV-----GAGTGIVI---DPNGVLTNNHVIAGATDINAFVSGSQTY 483
CC Db 147 TSQNNFTGTGGGSSSESGTGVFKDSDKAYITNNHVVGEANKLTV-TLYNGETE 205
CC QY 484 GVDVVGVDRTQDVAVLQRLRAGGLPSAAIGGG--VAVGEPVAMNGSGGGGTTPRAVPGR 541
CC Db 206 TAKLVGSDTTITDLAVLEISGNVKKVASFGDSQSLRTGKVKVIAIGNPLGQFSVTVQGI 265
CC QY 542 VALGQTVQASDSLGTGAETTLINGLIQPDAAIQPDGSGGPPVNGLGQVVGWNT 593
CC Db 266 ISGLNRTIDV-DTTQGTVM--NVLTQDAAINFGNSGGPLINASGOVIGINS 314
CC -----
CC RESULT 12
CC Y129_MYCTU
CC ID YY29_MYCTU STANDARD; PRT; 178 AA.

```


Db 41 VSVKVTQDKNHLHVLAKVSPVSTTRILLTSLFVNLFCNPSRYLSALALGD----- 93
QY 378 GLSGVLRVPRPVYVMPHSPHSPAGDIAPPALSDQRFADFPALPLDPS-AMVAQVGPQ----V 432
Db 94 -----PSVATVEDVSTV-----FPAGPLFPTTEGRIVQLFKERTYSV 130
QY 433 VNI-----NTKLGYNNAV-----GAGTGIVDPNGVVLTHNVVIAGATDIN----- 473
Db 131 VNIFFVTLRPLQKMTGVBIPEGNSGVVMDQGGYIVTNYHYVHGNALSRNPSFGDWGRV 190
QY 474 --AFSGSGQTVGVDDVGVYDRTQDVAVLQLRGAGGLPSAAI-----GGGVAVGEPVVA 524
Db 191 NILASGVQKNEKLVGADRAKDLAVKV-----DAPETLLPFIKVGQNSLKVGQOCLIA 246
QY 525 MNGSGGGGTTPRAVPGVRVVALGTVQVADSLGCAETLNGLLQFPAATQPGDSGFPVWG 584
Db 247 IGNPFGFDHT--LTVGVISGLNRDI---FSQTGV--TIGGGIQTDAINPGNSGGPLDLS 299
QY 585 LGQVGVGMNTA 594
Db 300 KGM LGIGINTA 309
RESULT 14
DEGS ECOLI
ID DEGS ECOLI STANDARD; PRT; 355 AA.
AC P31137; 1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease degs precursor (EC 3.4.21.-).
GN DEGS OR HHOB OR HHRH OR B3235 OR Z4594 OR ECS4108.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Bass S., Gu Q., Goddard A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=96165272; PubMed=8576051;
RA Waller P.R., Sauer R.T.;
RT "Characterization of degQ and degS, Escherichia coli genes encoding
RL homologs of the DegP protease."
RL J. Bacteriol. 178:1146-1153(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [5]

RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuwara S., Siba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
RN [6]
RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
RX MEDLINE=88105815; PubMed=3322223;
RA Vogel R.F., Entian K.-D., Mecke D.;
RT "Cloning and sequence of the mdh structural gene of Escherichia coli
coding for malate dehydrogenase."
RL Arch. Microbiol. 149:36-42(1987).
RN [7]
RP IDENTIFICATION
RA Bazan J.F., Fletterick R.J.;
RT "Structural and catalytic models of trypsin-like viral proteases."
RL Semin. Virol. 1:311-322(1990).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U15661; AAC43993.1; -;
DR EMBL; U32495; AAC44006.1; -;
DR EMBL; U18997; AAA58037.1; -;
DR EMBL; AE000402; AAC76267.1; -;
DR EMBL; AE005551; AAG58363.1; -;
DR EMBL; AP002564; BAB37531.1; -;
DR EMBL; M24777; -; NOT_ANNOTATED_CDS.
DR PIR; D91142; D91142.
DR MEROPS; S01.275; -;
DR EcoGene; EG11652; hohB.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_Ser_trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00105; PDZ; 1.
KW Hydrolase; Serine protease; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 355
FT DOMAIN 281 326
FT ACT_SITE 96 96
FT ACT_SITE 126 126
FT ACT_SITE 201 201
FT CONFLICT 253 253
FT CONFLICT 307 307
SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;
Query Match 7.5%; Score 225; DB 1; Length 355;
Best Local Similarity 30.9%; Pred.No. 7e-05;
Matches 67; Conservative 35; Mismatches 73; Indels 42; Gaps 9;
QY 396 PAAGDIAPPALSDQRFADFPALPLDPSAMVAQVGPVNTKLGYNNAVGA----- 447
Db 24 PSLRLSNP--LSTPQFDSTDETPASYNLAVRRAAPAVNV-----YNRGLNTNSHNLQLEI 76

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 15.5767 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796C-10

Perfect score: 3007

Sequence: 1 HMMHHHMMVDFGALPPEIN.....SGGPFVNGLGQVGMNTAAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1949	64.8	391	2 B70608	probable PPE prote
2	1656.5	55.1	396	2 H70741	probable PPE prote
3	1583	52.6	393	2 C70568	probable PPE prote
4	989	32.9	355	2 F70983	probable serine pr
5	779.5	25.9	393	2 G70929	probable PPE prote
6	778.5	25.9	361	2 S47170	hypothetical prote
7	754.5	25.1	354	2 A87242	probable secreted
8	753.5	25.1	409	2 A70932	probable PPE prote
9	737.5	24.5	403	2 H70931	probable PPE prote
10	737	24.5	423	2 B70931	probable PPE prote
11	731	24.3	421	2 H87056	PPE-family protein
12	705	23.4	408	2 G70925	probable PPE prote
13	702	23.3	391	2 B70625	probable PPE prote
14	702	23.3	413	2 F70560	probable PPE prote
15	690	22.9	468	2 B70932	probable PPE prote
16	688.5	22.9	463	2 C70931	probable PPE prote
17	675	22.4	380	2 A70646	probable PPE prote
18	668.5	22.2	394	2 G70881	probable PPE prote
19	667.5	22.2	385	2 H70503	probable PPE prote
20	635	21.1	350	2 H70929	probable PPE prote
21	633.5	21.1	365	2 F70929	probable PPE prote
22	615.5	20.5	402	2 A70882	probable PPE prote
23	606.5	20.2	423	2 C70582	probable PPE prote
24	602.5	20.0	391	2 D70922	probable PPE prote
25	595	19.8	394	2 A70504	probable PPE prote
26	589	19.6	406	2 F70675	probable PPE prote
27	575.5	19.1	391	2 A70663	probable PPE prote
28	521.5	17.3	3300	2 D70575	probable PPE prote
29	477.5	15.9	3716	2 B70969	probable PPE prote

ALIGNMENTS

RESULT 1

B70608

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: B70608

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: B70608

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-391 <COL>

A;Cross-references: GB:Z93777; GB:AL123456; NID:G3261726; PIDN:CAB07839.1; PID:e311073; I

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: PPE

Query Match 64.8%; Score 1949; DB 2; Length 391;

Best Local Similarity 100.0%; Pred. No. 1.3e-95;

Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	9	MVDFGALPPEIN	SARMVAGPGSASLVAAACMWDVSDLFSAASAFQSVVWGLTVGSWIG	68
Db	1	MVDFGALPPEIN	SARMVAGPGSASLVAAACMWDVSDLFSAASAFQSVVWGLTVGSWIG	60
Qy	69	SSAGLVAAASPYVAMSVTAGQAE	LTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI	128
Db	61	SSAGLVAAASPYVAMSVTAGQAE	LTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI	120
Qy	129	LIATNLGQNTPAIAVNEAEY	GEMWAQDAAPGYAAATATATATLLPPEAPEMTSAGG	188
Db	121	LIATNLGQNTPAIAVNEAEY	GEMWAQDAAPGYAAATATATATLLPPEAPEMTSAGG	180
Qy	189	LLEQAAVVEASDTAAANQ	LNNVPOALQQLAQPTCGTTPSSKLGGLWKTVPSPISN	248
Db	181	LLEQAAVVEASDTAAANQ	LNNVPOALQQLAQPTCGTTPSSKLGGLWKTVPSPISN	240
Qy	249	MVSMANNHSMNTNSGV	MTNTLSSMLKGFAPAAAQAVTAQNGVRAMSSLGSSG	308
Db	241	MVSMANNHSMNTNSGV	MTNTLSSMLKGFAPAAAQAVTAQNGVRAMSSLGSSG	300
Qy	309	LGSGVAANI	GRASVGLSVPOQAAANCAVTPAAEALPLTIS	368
Db	301	LGSGVAANI	GRASVGLSVPOQAAANCAVTPAAEALPLTIS	360
Qy	369	GQMGAPAGGL	SGVLRVPPRPVYMPHSPAAG	399

Db 361 GQMGARAGGSLGVLVRPPRPYVMPHSPAAG 391

RESULT 2

H70741

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70741

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; PMID:98295987; PMID:9634230

A: Accession: H70741

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-396 <COL>

A: Cross-references: GB:275555; GB:AL123456; NID:G3261608; PIDN:CAA99966.1; PID:e250360;

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 55.1%; Score 1656.5; DB 2; Length 396;

Best Local Similarity 85.1%; Pred. No. 3.1e-60;

Matches 336; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 9 MVDFGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSAAFSQSVVWGLTVGSGWIG 68

Db 1 MVDFGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSAAFSQSVVWGLTVGSGWIG 60

QY 69 SSAGLMVAASPYVAMSVTAGOAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128

Db 61 SSAGLMVAASPYVAMSVTAGOAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

QY 129 LIATNLGQNTPAIVNAEYGEWQAQDAAMFGYAAATATATATLPPPEAPEMTSAG 188

Db 121 LIATNLGQNTPAIVNAEYGEWQAQDAAMFGYAAATATATLPPPEAPEMTSAG 180

QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGTTSSKGLGKLVKTVSPHRSPIIN 248

Db 181 LLEQAAVEEADITAAANQLMNNVPQALQQAQPTGTTSSKGLGKLVKTVSPHRSPIIN 240

QY 249 MVSNNHNSMTNSGVSMNTLSMLKGFAPAAAQAVOTAQNGVRAMSS---LGSSL 304

Db 241 IVSMLNNHNSMTNSGVSMNTLSMLKGFAP-AAAQAVETAQNGVQVMSLSLQSSLS 299

QY 305 GSSGLGGVAANLGRAASVGSLSVPPQAWAANAQAVTPAARALPLTSLTSAARPGQMLG 364

Db 300 GSSGLGGVAANLGRAASVGSLSVPPQAWAANAQAVTPAARALPLTSLTSAARPGHMLG 359

QY 365 GLPVGMQARAG--GGLSGVLVRPPRPYVMPHSPAAG 399

Db 360 GLPLGQLTNSGGGFGVSNALRMPPRAYVMPRPAAG 396

RESULT 3

C70568

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: C70568

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; PMID:98295987; PMID:9634230

A: Accession: C70568

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-393 <COL>

A: Cross-references: GB:295390; GB:AL123456; NID:G3261766; PIDN:CAB08702.1; PID:e316074;

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 52.6%; Score 1583; DB 2; Length 393;

Best Local Similarity 81.5%; Pred. No. 2.3e-76;

Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 9 MVDFGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSAAFSQSVVWGLTVGSGWIG 68

Db 1 MVDFGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSAAFSQSVVWGLTVGSGWIG 60

QY 69 SSAGLMVAASPYVAMSVTAGOAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128

Db 61 SSAGLMVAASPYVAMSVTAGOAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMT 120

QY 129 LIATNLGQNTPAIVNAEYGEWQAQDAAMFGYAAATATATLPPPEAPEMTSAG 188

Db 121 LIATNLGQNTPAIVNAEYGEWQAQDAAMFGYAAATATATLPPPEAPEMTSAG 180

QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGTTSSKGLGKLVKTVSPHRSPIIN 248

Db 181 LLEQAAVEEADITAAANQLMNNVPQALQQAQPTGTTSSKGLGKLVKTVSPHRSPIIN 240

QY 249 MVSNNHNSMTNSGVSMNTLSMLKGFAPAAAQAVOTAQNGVRAMSS---LGSSL 304

Db 241 VSSANNHNSMTNSGVSMNTLSMLKGFAP-AAAQAVETAQNGVQVMSLSLQSSLS 299

QY 305 GSSGLGGVAANLGRAASVGSLSVPPQAWAANAQAVTPAARALPLTSLTSAARPGQMLG 364

Db 300 GSSGLGGVAANLGRAASVGSLSVPPQAWAANAQAVTPAARALPLTSLTSAARPGHMLG 359

QY 365 GLPVGMQARAGGGLSGVLVRPPRPYVMPHSPAAG 399

Db 360 GLPLGHT-SVNAGSINNLRVPAAYAIPTPAAG 393

RESULT 4

F70983

probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C:Accession: F70983

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; PMID:98295987; PMID:9634230

A: Accession: F70983

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-355 <COL>

A: Cross-references: GB:296071; GB:AL123456; NID:G3242254; PIDN:CAB09453.1; PID:G2181967

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: pepA

C: Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp:

Query Match 32.9%; Score 989; DB 2; Length 355;

Best Local Similarity 98.0%; Pred. No. 3.4e-45;

Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 397 AAGDIAPALSDQDFADFPALPLDPSAMVAGVQVQVNVINTKLYNNVAVGAGTGVIDPN 456

Db 28 APAQAAPALSDQDFADFPALPLDPSAMVAGVQVQVNVINTKLYNNVAVGAGTGVIDPN 87

QY 457 GVVLTNNHVIAGATDINAFSVGSGQTTGVVDVGVDRQDVAVLQRCAGGLPSAIGGV 516

RESULT 8

A70932

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: A70932

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 9825987; PMID: 9634230

A: Accession: A70932

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-409 <COL>

A: Cross-references: GB:AL022021; GB:AL123456; NID: g3250699; PIDN: CAA17729.1; PID: e125461

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 25.1%; Score 753.5; DB 2; Length 409;
 Best Local Similarity 42.6%; Pred. No. 9.8e-33;
 Matches 182; Conservative 54; Mismatches 152; Indels 39; Gaps 11;

QY 10 VDFGALPPEINSARMYAGPGSASIVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWS 69

DB 1 MDFGALPPEINSARMYAGPGSGGLLAAADWDLAAELYSAAASYGSTEIGLTVAPWGP 60

QY 70 SAGLMVAAASPYVAMSVTGAQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMIL 129

DB 61 SSITMAAAVAPYVAMISVTGAQEAQAKIAAGVYETAFAATVPPVPIEARNALLMSL 120

QY 130 IATNLGONTPTAIVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPMTSAGGL 189

DB 121 VAINIFONTPTAATEAHYAENWAOAAAMYGAGSATA-SQLAPPSEPPPTNPST 179

QY 190 LEQAAVVEEASDTAAA-----NQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPH 242

DB 180 AAQSAVVAQAAGAAASDDITAQLSQLISLPLTQLSLA--TTATATASAG--WDIV--- 232

QY 243 RSPISNMVSMNMHMTNSGVS-----MTNTLSSMLKGFAPAAAQAVQTAQNGVRAM 297

DB 233 LQSIITTLANTGPYSIIGLAI PGGMWLTFGQILGLAQNAPGVAAALGPKAAAGALSPL 292

QY 298 SSL-GSSLGS-SGLGGVAAANLGRAASVGLSVFQAWAANQAVTPAARALPLTSLTA- 354

DB 293 APLRGYIGDITPLGGATGIAIAIVGSLSVPGWAEAAAPVRAVASVLPTGAPAL 352

QY 355 AERPGQMLGLPVQNGARAGGL-----SGVLRVPPRPVPHSPAPAGDIAPPALSD 409

DB 353 AAAPAGALFGEMALSSLAGALAGTAVRSAGAARV-----AGGSVTEVDASTT 401

QY 410 RFADFP 416

DB 402 TIIVIPA 408

RESULT 9

H70931

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70931

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 9825987; PMID: 9634230

A: Accession: H70931

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-403 <COL>

A: Cross-references: GB:AL022021; GB:AL123456; NID: g3250699; PIDN: CAA17728.1; PID: e125461

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 24.5%; Score 737.5; DB 2; Length 403;
 Best Local Similarity 42.6%; Pred. No. 6.7e-32;
 Matches 181; Conservative 60; Mismatches 123; Indels 61; Gaps 12;

QY 10 VDFGALPPEINSARMYAGPGSASIVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWS 69

DB 5 LDFATLPEINSARMYSGAGSAPMLAAASAHGLSAELRASALSYSVLSLTGEEWHP 64

QY 70 SAGLMVAAASPYVAMSVTGAQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMIL 129

DB 65 ASASMTAAAPYVAMSVTAVRAEQAGQAQAEAAAAYEAAFAATVPPVIEANRAQLMAL 124

QY 130 IATNLGONTPTAIVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPMTSAGGL 189

DB 125 IATNLGONAPAIATTAQAIAEWSQDAAMYGAGSAAAT-QLTFTEPVTNAGSL 183

QY 190 LEQAAVVEEASDTAAA-----NQLMNNVPOALQQLAQTQGT-----TPSS 230

DB 184 AAQSAIAHAATGASAGAQOTTLSQIAAIPSVLQGLSSSTAATFASGPSGLLGVSGSS 243

QY 231 KLGLWKTVPSPHSPISNMVSMNMHMTNSGYSM-TNTLS---SMLKGFAPAAAQAV 286

DB 244 WLDKLMALDLPN-----SNFWNTIASSGGLFPLSNTIAPPLGLGGVAAADAAGDV 293

QY 287 QTAQNGVRAMSSLGSSL-----GSSGLGGVAAANLGRAASVGLSVFQAWAANQAVTP 341

DB 294 LGATSG-----GLGGALVAPLGSAGGLGTVAGLGNAAATVGLTSPSWTAAAPLSP 348

QY 342 AARAL---PLTSLTSAERGPQMLGGLPVQNGARAGGLSGVLRVPP-----RPYMPH 394

DB 349 LGSALGCTPMVAPPVPAAG---MPGPFMTMGQGGFG-----RAVPOYGRPRNFVAR 398

QY 395 SPAAG 399

DB 399 PPAAG 403

RESULT 10

B70931

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70931

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 9825987; PMID: 9634230

A: Accession: B70931

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-423 <COL>

A: Cross-references: GB:AL022021; GB:AL123456; NID: g3250699; PIDN: CAA17722.1; PID: e125461

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 24.5%; Score 737; DB 2; Length 423;
 Best Local Similarity 41.8%; Pred. No. 7.6e-32;
 Matches 183; Conservative 50; Mismatches 125; Indels 80; Gaps 12;

QY 10 VDFGALPPEINSARMYAGPGSASIVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSGWS 69

Db 1 MDFGLPPPEINSGRMTYGPGLMLAAATWDLGAVELHATAAGVASELSALT-GAWSGP 59
QY 70 SAGLWVAASPYVAMSVTAGAELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 129
Db 60 SSTMSASAAPYVAMSVATAVHAELAGACARLAIAYEAFATVPPPPVIAENRAELMIL 119
QY 130 IATNLLGONTPTAIVNEAYGEMWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGL 189
Db 120 IATNIFGONTPTAIVNEAYGEMWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGL 178
QY 190 LEQAAAVEASDTAAAN-----OLMNVPOALQLOAQT-----QGTFP-----S 229
Db 179 GQSSAVACTAATAGGNLQSAFPQLLSAVPALQGLALPTASQASATPQWVTDGLNLS 238
QY 230 SKLGGJWKTVPSPHRSPISNMVSVANNHMTNSGVSMNTLSSMLKGFAPAAAAQVTA 289
Db 239 TELGG-AVTGPYTFP-----GVLPPSGVPYLLIGQSVL-----V 271
QY 290 AONGVRAVSSLSG-----SIGSSGLGGG-VAANLGRAASVGSLS 327
Db 272 TONGQGVSAALLKIGKPKITGALAPLAEFALHTPILGSEGLGGVSVAGIGRAGLVGKLS 331
QY 328 VPQWAAANQAVTPAARALELTLTS---AABERGQOMLGGLPVQGMGARAGGLSGVL 384
Db 332 VPQGTVAARPEIPSPAALQATRLAAAPTAAATDAGALLGGVSLGSLAGRAAGSTG--- 388
QY 385 VPPRPVPMHSPAAAGDIA 402
Db 389 ---HPTGSAAPAVGAAA 403

RESULT 11
H87056
PPE-family protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87056
R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hoo
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H87056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL450380; NID:gl3093150; PIDN:CAC31563.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML1182

Query Match 24.3%; Score 731; DB 2; Length 421;
Best Local Similarity 41.1%; Pred. No. 1.6e-31;
Matches 174; Conservative 55; Mismatches 160; Indels 34; Gaps 8;

QY 9 MVDGALPPEINSGRMTYGPGLMLAAATWDLGAVELHATAAGVASELSALT-GAWSGP 68
Db 1 MFDFAALSPEINSGRMTYGPGLMLAAATWDLGAVELHATAAGVASELSALT-GAWSGP 59
QY 69 SAGLWVAASPYVAMSVTAGAELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 128
Db 60 ESAAALAEVTPYKELTQNAASAEALTATQTLVAANAYETATMTVPPPLMVFVNRQAQL 119
QY 129 LIATNLLGONTPTAIVNEAYGEMWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGL 188
Db 120 LINSNTFGQNTAIVNEAYGEMWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGL 179
QY 189 L-----LEQAAAVEASDTAAANQALMN-----VPQALQ-----LAQP 222
Db 180 LAQVEEVVEEVVEEVVEEAEQAISQAALDQAVNEGMEATVVPQVDQVQVNDVATP 239

QY 223 TQGTTPSSKLG--LWKTVPSPHRSPISNMVSVANNHMTNSGVSMNTLSSMLKGFAPA 280
Db 240 QTAVPDSSAAAAAQQWGFAGHLSPTLSTNNHAGMANAGLSVNGSGMSAKSLAP- 298
QY 281 AAAQAVTAAQNGVRAMSSLSGSLGGVAAANLGRAASVGSLSVPOWAAANQAVT 340
Db 299 TTTTAAESAFAKMSAVQSTGRGLGSSGGHVTQAOLGRAASIGSLVFPQTWTTASQPT 358
QY 341 PAARALPLTSLTSAERGPQOML-GGLPVQGM---GARAGGLSGVLURVPPRPVPMHSP 396
Db 359 AATRALSPARVAVATESAPLGGGLPMAPVPGGSGTGGVNTALRLQFRAFVMPRNP 418
QY 397 AAG 399
Db 419 AAG 421

RESULT 12
G70925
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70925
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70925
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-408 <COL>
A:Cross-references: GB:Z74024; GB:AL123456; NID:g3250700; PIDN:CAA98377.1; PID:el301025;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 23.4%; Score 705; DB 2; Length 408;
Best Local Similarity 41.8%; Pred. No. 3.5e-30;
Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;
QY 10 VDFGALPPEINSGRMTYGPGLMLAAATWDLGAVELHATAAGVASELSALT-GAWSGP 69
Db 1 MDFGVLPEINSGRMTYGPGLMLAAATWDLGAVELHATAAGVASELSALT-GAWSGP 60
QY 70 SAGLWVAASPYVAMSVTAGAELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 129
Db 61 AAASMWAAVTPYVAVLSATAGQAQAGVQARAAAAAYELAFAMTVPPPVVAVNALLVAL 120
QY 130 IATNLLGONTPTAIVNEAYGEMWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGL 189
Db 121 VATNFFGONTPTAIVNEAYGEMWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGL 179
QY 190 LEQAAA-----VEASDTAAANQALMN-----VP--QALQ--LAQGTQGTTPSSKL--- 232
Db 180 AGQAAATVSTVPPPLATTAAVPLQQLSSTSLIPWYSAALQWLAENLLGLTPNRRITV 239
QY 233 -----GGLWKTVPSPHRSPISNMVSVANNHMTNSGVSMNTLSSMLKGFAPAAAA 283
Db 240 RLIGISVDFEGL-----LOFEASLAQQAIPGTPGGAG--DSGSSVLDSHGPTIFA 287
QY 284 QAVQTAAQNGVRAMSSLSGSLGS-----SGLGGVAAANLGRAASVGSLS 327
Db 288 -----GPRASPSVAGGAVGVQTPQPYWYVWALDRESIGGSVAALKGSSAGLS 338
QY 328 VPQWAAANQAVTAAARALP---LTSLSAERGPQOMLGGLPVQGMGARAGGLSGVL 384
Db 339 VPPDWAARAWANPAANWLPFGDDVTALRGAENA---LLRGFPVASAGQSTGGF---VHK 393
QY 385 VPPRPVPMHSPAAAG 399

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:52:58 ; Search time 46.4747 Seconds
(without alignments)
3625.462 Million cell updates/sec

Title: US-09-597-796C-10

Perfect score: 3007
Sequence: 1 HMHHEHHMVDGALPPEIN.....SGGPVYVGLGVGMNTAAS 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3007	100.0	596	9	US-09-287-849-26
2	3007	100.0	596	12	US-09-886-349A-20
3	3007	100.0	596	14	US-10-359-460-26
4	3007	100.0	596	15	US-10-098-732A-20
5	2946	98.0	729	14	US-10-369-983-21
6	2946	98.0	813	15	US-10-369-983-15
7	2946	98.0	825	15	US-10-369-983-14
8	2946	98.0	875	15	US-10-369-983-13
9	2946	98.0	930	14	US-10-098-732A-65
10	2946	98.0	930	15	US-10-369-983-12
11	2946	98.0	1016	15	US-10-369-983-18
12	2946	98.0	1022	15	US-10-369-983-17
13	2946	98.0	1154	15	US-10-369-983-16
14	2943	97.9	723	12	US-09-886-349A-18
15	2943	97.9	723	14	US-10-098-732A-18

16	2943	97.9	729	15	US-10-369-983-22
17	2941	97.8	729	9	US-09-287-849-2
18	2941	97.8	729	12	US-09-886-349A-16
19	2941	97.8	729	14	US-10-359-460-2
20	2941	97.8	723	14	US-10-098-732A-16
21	2941	97.8	729	15	US-10-359-459-2
22	2033.5	67.6	600	9	US-09-287-849-22
23	2033.5	67.6	600	14	US-10-359-460-22
24	1955	65.0	1010	15	US-10-369-983-4
25	1949	64.8	391	12	US-09-886-349A-14
26	1949	64.8	391	14	US-10-193-002-102
27	1949	64.8	391	14	US-10-084-843-107
28	1949	64.8	391	14	US-10-098-732A-14
29	1949	64.8	723	15	US-10-369-983-2
30	1934	64.3	391	12	US-09-872-186-8
31	1652.5	55.0	396	14	US-10-193-002-106
32	1652.5	55.0	396	14	US-10-084-843-111
33	1583	52.6	393	12	US-10-282-122A-62455
34	1583	52.6	393	12	US-10-282-122A-64892
35	1486.5	49.4	359	14	US-10-193-002-104
36	1486.5	49.4	359	14	US-10-084-843-109
37	1189	39.5	358	9	US-09-287-849-8
38	1189	39.5	358	14	US-10-359-460-8
39	1187	39.5	263	12	US-09-886-349A-12
40	1187	39.5	263	14	US-10-193-002-92
41	1187	39.5	263	14	US-10-084-843-91
42	1187	39.5	263	14	US-10-098-732A-12
43	989	32.9	355	9	US-09-712-363-161
44	987	32.8	330	12	US-09-886-349A-4
45	987	32.8	330	14	US-10-098-732A-4

ALIGNMENTS

RESULT 1

US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 100.0%; Score 3007; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;

Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMHGHHHMDVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60
DB 1 HMHGHHHMDVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWIGSSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIA 120
DB 61 LTVGSWIGSSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIA 120

QY 121 ENRAELMLIATNLGONTPAIANNVNEAYGEMWQADAAAFGYAAATATATATALLPEEA 180
DB 121 ENRAELMLIATNLGONTPAIANNVNEAYGEMWQADAAAFGYAAATATATATALLPEEA 180

QY 181 PEMSAGLLLEQAAVEASDPTAAANQLMNNVPQALQOQLAQPTGTTTPSSKLGGLWKTVS 240
DB 181 PEMSAGLLLEQAAVEASDPTAAANQLMNNVPQALQOQLAQPTGTTTPSSKLGGLWKTVS 240

QY 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAACNGVRAMSSL 300
DB 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAACNGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGPG 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGPG 360

QY 361 QMLGGLPVQMGARAGGGLSGVLVPPRPVYMPHSPAAAGDIAPPALSDQDFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGGLSGVLVPPRPVYMPHSPAAAGDIAPPALSDQDFADFPALPLD 420

QY 421 PSAMVAQVGPVQVNTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPVQVNTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480

QY 481 QTYGVVGVYDRTQDVAVLQRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPG 540
DB 481 QTYGVVGVYDRTQDVAVLQRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTAETLNGLIQFDDAAIQFGDSGGPVVNGLGQVGMNTAAS 596
DB 541 RVVALGQTVQASDSLTAETLNGLIQFDDAAIQFGDSGGPVVNGLGQVGMNTAAS 596

RESULT 2
US-09-886-349A-20
; Sequence 20, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TB9-Ra35 (Designated MTB59F)
US-09-886-349A-20

Query Match 100.0%; Score 3007; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;

Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMHGHHHMDVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60
DB 1 HMHGHHHMDVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWIGSSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIA 120
DB 61 LTVGSWIGSSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIA 120

QY 121 ENRAELMLIATNLGONTPAIANNVNEAYGEMWQADAAAFGYAAATATATATALLPEEA 180
DB 121 ENRAELMLIATNLGONTPAIANNVNEAYGEMWQADAAAFGYAAATATATATALLPEEA 180

QY 181 PEMSAGLLLEQAAVEASDPTAAANQLMNNVPQALQOQLAQPTGTTTPSSKLGGLWKTVS 240
DB 181 PEMSAGLLLEQAAVEASDPTAAANQLMNNVPQALQOQLAQPTGTTTPSSKLGGLWKTVS 240

QY 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAACNGVRAMSSL 300
DB 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAACNGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGPG 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAAANQAVTPAARALPLTSLTSAABERGPG 360

QY 361 QMLGGLPVQMGARAGGGLSGVLVPPRPVYMPHSPAAAGDIAPPALSDQDFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGGLSGVLVPPRPVYMPHSPAAAGDIAPPALSDQDFADFPALPLD 420

QY 421 PSAMVAQVGPVQVNTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPVQVNTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480

QY 481 QTYGVVGVYDRTQDVAVLQRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPG 540
DB 481 QTYGVVGVYDRTQDVAVLQRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTAETLNGLIQFDDAAIQFGDSGGPVVNGLGQVGMNTAAS 596
DB 541 RVVALGQTVQASDSLTAETLNGLIQFDDAAIQFGDSGGPVVNGLGQVGMNTAAS 596

RESULT 3
US-10-359-460-26
; Sequence 26, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT FILING DATE: 2003-02-05
; PRIOR FILING DATE: 2003-02-05
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-26
```

```
Query Match      100.0%; Score 3007; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMEHHHHHWDVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60
DB 1 HMEHHHHHWDVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSGWIGSSAGLMVAASPYVAMSVTAGQAELETAQVRVAAAYETAYGLTVPFPVIA 120
DB 61 LTVGSGWIGSSAGLMVAASPYVAMSVTAGQAELETAQVRVAAAYETAYGLTVPFPVIA 120

QY 121 ENRAELMILIAATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEA 180
DB 121 ENRAELMILIAATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEA 180

QY 181 PEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQQLAQTGCTTSSKLGGLWKTYS 240
DB 181 PEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQQLAQTGCTTSSKLGGLWKTYS 240

QY 241 PHRSPISNNVMANNHMTNSGVSMTNTLSMLKGFAPAAAQVQTAAGNGVRAMSSL 300
DB 241 PHRSPISNNVMANNHMTNSGVSMTNTLSMLKGFAPAAAQVQTAAGNGVRAMSSL 300

QY 301 GSSLGSGGLGGGVAANLGRAASVGSLSVPOAWAANQVTPAARALPLTSLTSAERGPG 360
DB 301 GSSLGSGGLGGGVAANLGRAASVGSLSVPOAWAANQVTPAARALPLTSLTSAERGPG 360

QY 361 QMLGGLPVQMGARAGGGLSGVLRVPPRYVPHSPFAAGDIAPPALSDRFPALPLD 420
DB 361 QMLGGLPVQMGARAGGGLSGVLRVPPRYVPHSPFAAGDIAPPALSDRFPALPLD 420

QY 421 PSAMVAQVGPQVYNTKLGYNNAVAGTGIVDPNGVVLTNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVYNTKLGYNNAVAGTGIVDPNGVVLTNHVIAGATDINAFSVGSG 480

QY 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGCGVAVGEPVAMGNSGGGQGTTPRAVPG 540
DB 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGCGVAVGEPVAMGNSGGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAAS 596
DB 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAAS 596
```

```
RESULT 4
US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Braumon, Mark
; APPLICANT: Guderman, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US/10/098,732A
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TDH9-Ra35 (designated MTB59F)
US-10-098-732A-20
```

```
Query Match      100.0%; Score 3007; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMEHHHHHWDVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60
DB 1 HMEHHHHHWDVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSGWIGSSAGLMVAASPYVAMSVTAGQAELETAQVRVAAAYETAYGLTVPFPVIA 120
DB 61 LTVGSGWIGSSAGLMVAASPYVAMSVTAGQAELETAQVRVAAAYETAYGLTVPFPVIA 120

QY 121 ENRAELMILIAATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEA 180
DB 121 ENRAELMILIAATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEA 180

QY 181 PEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQQLAQTGCTTSSKLGGLWKTYS 240
DB 181 PEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQQLAQTGCTTSSKLGGLWKTYS 240

QY 241 PHRSPISNNVMANNHMTNSGVSMTNTLSMLKGFAPAAAQVQTAAGNGVRAMSSL 300
DB 241 PHRSPISNNVMANNHMTNSGVSMTNTLSMLKGFAPAAAQVQTAAGNGVRAMSSL 300

QY 301 GSSLGSGGLGGGVAANLGRAASVGSLSVPOAWAANQVTPAARALPLTSLTSAERGPG 360
DB 301 GSSLGSGGLGGGVAANLGRAASVGSLSVPOAWAANQVTPAARALPLTSLTSAERGPG 360

QY 361 QMLGGLPVQMGARAGGGLSGVLRVPPRYVPHSPFAAGDIAPPALSDRFPALPLD 420
DB 361 QMLGGLPVQMGARAGGGLSGVLRVPPRYVPHSPFAAGDIAPPALSDRFPALPLD 420

QY 421 PSAMVAQVGPQVYNTKLGYNNAVAGTGIVDPNGVVLTNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVYNTKLGYNNAVAGTGIVDPNGVVLTNHVIAGATDINAFSVGSG 480

QY 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGCGVAVGEPVAMGNSGGGQGTTPRAVPG 540
DB 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGCGVAVGEPVAMGNSGGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAAS 596
DB 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAAS 596
```

```
RESULT 5
US-10-369-983-21
; Sequence 21, Application US/10369983
; Publication No. US20030235591A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderman, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:MTB72F

US-10-369-983-21

Query Match 98.0%; Score 2946; DB 15; Length 729;
Best Local Similarity 100.0%; Pred. No. 1.2e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
DB 142 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201

QY 69 SSAGLWVAASPPYVWMSVTTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128
DB 202 SSAGLWVAASPPYVWMSVTTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261

QY 129 LIATNLLGONTPTAIVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
DB 262 LIATNLLGONTPTAIVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 321

QY 189 LLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248
DB 322 LLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 381

QY 249 MVSMAHHMNTSGVSMNTTLLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSLGSSG 308
DB 382 MVSMAHHMNTSGVSMNTTLLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSLGSSG 441

QY 309 LGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 442 LGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 501

QY 369 GQMGARAGGGLSGVLRVPPRPYPWPHSPAAGDIAPPALSDQDFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPPRPYPWPHSPAAGDIAPPALSDQDFADFPALPLDPSAMVAQV 561

QY 429 GPQVNTIKLGYNNVAGAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVV 488
DB 562 GPQVNTIKLGYNNVAGAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVV 621

QY 489 GYDRTQDVAVLQIRGAGGLPSAIIQGGVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQT 548
DB 622 GYDRTQDVAVLQIRGAGGLPSAIIQGGVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQT 681

QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGPPVNGLGQVVGMMNTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGPPVNGLGQVVGMMNTAAS 729

RESULT 6

US-10-369-983-15
Sequence 15, Application US/10369983
Publication No. US20030235593A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
CURRENT FILING DATE: 2003-02-18
PRIOR APPLICATION NUMBER: US 60/357,351
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 15
LENGTH: 813
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:fusion protein
OTHER INFORMATION: MTB81F (MTB72F-DPV)

US-10-369-983-15

Query Match 98.0%; Score 2946; DB 15; Length 813;
Best Local Similarity 100.0%; Pred. No. 1.4e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
DB 142 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201

QY 69 SSAGLWVAASPPYVWMSVTTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128
DB 202 SSAGLWVAASPPYVWMSVTTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261

QY 129 LIATNLLGONTPTAIVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
DB 262 LIATNLLGONTPTAIVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 321

QY 189 LLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248
DB 322 LLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 381

QY 249 MVSMAHHMNTSGVSMNTTLLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSLGSSG 308
DB 382 MVSMAHHMNTSGVSMNTTLLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSLGSSG 441

QY 309 LGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 442 LGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 501

QY 369 GQMGARAGGGLSGVLRVPPRPYPWPHSPAAGDIAPPALSDQDFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPPRPYPWPHSPAAGDIAPPALSDQDFADFPALPLDPSAMVAQV 561

QY 429 GPQVNTIKLGYNNVAGAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVV 488
DB 562 GPQVNTIKLGYNNVAGAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVV 621

QY 489 GYDRTQDVAVLQIRGAGGLPSAIIQGGVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQT 548
DB 622 GYDRTQDVAVLQIRGAGGLPSAIIQGGVAVGEPVWAMGNSGGGGTTPRAVGRVVALGQT 681

QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGPPVNGLGQVVGMMNTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGPPVNGLGQVVGMMNTAAS 729

RESULT 7

US-10-369-983-14
Sequence 14, Application US/10369983
Publication No. US20030235593A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
CURRENT FILING DATE: 2003-02-18
PRIOR APPLICATION NUMBER: US 60/357,351
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 825
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:fusion protein
OTHER INFORMATION: MTB83F (MTB72F-MT1)

US-10-369-983-14

Query Match 98.0%; Score 2946; DB 15; Length 825;
Best Local Similarity 100.0%; Pred. No. 1.4e-137;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MYDFGALPEINSAARMYAGPGSASLVAAQAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 68
DB 142 MYDFGALPEINSAARMYAGPGSASLVAAQAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 201
QY 69 SSAGLVAAASPYVAVWMSVTAGQAELETAQVRVAAAAYETAYGLTPPPVIAENRAELMI 128
DB 202 SSAGLVAAASPYVAVWMSVTAGQAELETAQVRVAAAAYETAYGLTPPPVIAENRAELMI 261
QY 129 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 188
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 321
QY 189 LLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248
DB 322 LLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381
QY 249 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLSGSSG 308
DB 382 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLSGSSG 441
QY 309 LGGVAAANLGRAASVGSLSVPOQAAAAQAQVTAQAQNGVRAMSSLSGSSG 368
DB 442 LGGVAAANLGRAASVGSLSVPOQAAAAQAQVTAQAQNGVRAMSSLSGSSG 501
QY 369 GQMGARAGGGLSGVLRVPPRYVMPHSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPPRYVMPHSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNVINTKLYNNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 488
DB 562 GPQVNVINTKLYNNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 621
QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEVPVAMGNSGGQGTTPRAVGRVVALGQT 548
DB 622 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEVPVAMGNSGGQGTTPRAVGRVVALGQT 681
QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAAS 729

RESULT 8
US-10-369-983-13
; Sequence 13, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Gudarian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 875
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-Erd14)
US-10-369-983-13

Query Match 98.0%; Score 2946; DB 15; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.5e-137;

Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MYDFGALPEINSAARMYAGPGSASLVAAQAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 68
DB 142 MYDFGALPEINSAARMYAGPGSASLVAAQAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 201
QY 69 SSAGLVAAASPYVAVWMSVTAGQAELETAQVRVAAAAYETAYGLTPPPVIAENRAELMI 128
DB 202 SSAGLVAAASPYVAVWMSVTAGQAELETAQVRVAAAAYETAYGLTPPPVIAENRAELMI 261
QY 129 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 188
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 321
QY 189 LLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248
DB 322 LLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381
QY 249 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLSGSSG 308
DB 382 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLSGSSG 441
QY 309 LGGVAAANLGRAASVGSLSVPOQAAAAQAQVTAQAQNGVRAMSSLSGSSG 368
DB 442 LGGVAAANLGRAASVGSLSVPOQAAAAQAQVTAQAQNGVRAMSSLSGSSG 501
QY 369 GQMGARAGGGLSGVLRVPPRYVMPHSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPPRYVMPHSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNVINTKLYNNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 488
DB 562 GPQVNVINTKLYNNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 621
QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEVPVAMGNSGGQGTTPRAVGRVVALGQT 548
DB 622 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEVPVAMGNSGGQGTTPRAVGRVVALGQT 681
QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAAS 729

RESULT 9
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Gudarian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72F-MAPS
; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Rai12-TbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-65

Query Match 98.0%; Score 2946; DB 14; Length 930;

Best Local Similarity 100.0%; Pred. No. 1.6e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 68
DB 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 201
QY 69 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
DB 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
QY 129 LIATNLGONTPALAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPFEAPMTSAGG 188
DB 262 LIATNLGONTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 321
QY 189 LLEQAAAVEEASDTAAANQLMNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHRSPISN 248
DB 322 LLEQAAAVEEASDTAAANQLMNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHRSPISN 381
QY 249 MVSMAHHMNTSGVSMVNTLSSMLKGFAPAAAAQAVTAAQNGVRAVSSLSLSSGSSG 308
DB 382 MVSMAHHMNTSGVSMVNTLSSMLKGFAPAAAAQAVTAAQNGVRAVSSLSLSSGSSG 441
QY 309 LGGGVAANLGRAASVGSLSVPOQAAAAQAVTAAQNGVRAVSSLSLSSGSSG 368
DB 442 LGGGVAANLGRAASVGSLSVPOQAAAAQAVTAAQNGVRAVSSLSLSSGSSG 501
QY 369 GOMGARAGGSLGVLRPPRYVPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GOMGARAGGSLGVLRPPRYVPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNTNTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTGYGVDVV 488
DB 562 GPQVNTNTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTGYGVDVV 621
QY 489 GYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVAVMNGSGGGGTTPRAVGRVVALGOT 548
DB 622 GYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVAVMNGSGGGGTTPRAVGRVVALGOT 681
QY 549 VQASDSLTGAETLNGLIQFDDAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 682 VQASDSLTGAETLNGLIQFDDAIQPGDSGGPVVNGLGQVVGMMNTAAS 729

RESULT 10

US-10-369-983-12
; Sequence 12, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: R95F (MTB72F-NAPs)
US-10-369-983-12

Query Match 98.0%; Score 2946; DB 15; Length 930;
Best Local Similarity 100.0%; Pred. No. 1.6e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 68
DB 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 201
QY 69 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
DB 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
QY 129 LIATNLGONTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 188
DB 262 LIATNLGONTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 321
QY 189 LLEQAAAVEEASDTAAANQLMNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHRSPISN 248
DB 322 LLEQAAAVEEASDTAAANQLMNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHRSPISN 381
QY 249 MVSMAHHMNTSGVSMVNTLSSMLKGFAPAAAAQAVTAAQNGVRAVSSLSLSSGSSG 308
DB 382 MVSMAHHMNTSGVSMVNTLSSMLKGFAPAAAAQAVTAAQNGVRAVSSLSLSSGSSG 441
QY 309 LGGGVAANLGRAASVGSLSVPOQAAAAQAVTAAQNGVRAVSSLSLSSGSSG 368
DB 442 LGGGVAANLGRAASVGSLSVPOQAAAAQAVTAAQNGVRAVSSLSLSSGSSG 501
QY 369 GOMGARAGGSLGVLRPPRYVPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GOMGARAGGSLGVLRPPRYVPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNTNTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTGYGVDVV 488
DB 562 GPQVNTNTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVSGGTGYGVDVV 621
QY 489 GYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVAVMNGSGGGGTTPRAVGRVVALGOT 548
DB 622 GYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVAVMNGSGGGGTTPRAVGRVVALGOT 681
QY 549 VQASDSLTGAETLNGLIQFDDAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 682 VQASDSLTGAETLNGLIQFDDAIQPGDSGGPVVNGLGQVVGMMNTAAS 729

RESULT 11

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 98.0%; Score 2946; DB 15; Length 1016;
Best Local Similarity 100.0%; Pred. No. 1.8e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 68

Db	142	MDVFGALPPEIISARIYAGPGSASLVAAQMWDSVADLSAASAFQSVVWGLTGVGSWIG	201
Qy	69	SSAGLMAAASPYVAWMSVTAGCAELTAAQVYVAAAAYETAYGLTVPVPPIAENRAELMI	128
Db	202	SSAGLMAAASPYVAWMSVTAGCAELTAAQVYVAAAAYETAYGLTVPVPPIAENRAELMI	261
Qy	129	LIATNLLGONTTFAIAVNEAEYGEWMAQDAAMFGYAAATATATATATALLPFEAPEMTSAGS	188
Db	262	LIATNLLGONTTFAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGS	321
Qy	189	LLEQAAVEEASDTAAANQLMNNVPALQQLAQTQGTTPSSKLGGLWKTVSPHRSPISN	248
Db	322	LLEQAAVEEASDTAAANQLMNNVPALQQLAQTQGTTPSSKLGGLWKTVSPHRSPISN	381
Qy	249	MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLSGSSLGSSG	308
Db	382	MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLSGSSLGSSG	441
Qy	309	LGGSVAAANLGRAASVGSLSVQAAWAAANQVTPPAARALPTSTLSAABERGPGOMLGLBV	368
Db	442	LGGSVAAANLGRAASVGSLSVQAAWAAANQVTPPAARALPTSTLSAABERGPGOMLGLBV	501
Qy	369	GQMGARAGGSLGVLVPPREPYMHPHSPAAGDIAPPALSQDRPADFPALPLDPSAMVAQV	428
Db	502	GQMGARAGGSLGVLVPPREPYMHPHSPAAGDIAPPALSQDRPADFPALPLDPSAMVAQV	561
Qy	429	GPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTNHHVIGATDINAFSVSGSGQTYGVGV	488
Db	562	GPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTNHHVIGATDINAFSVSGSGQTYGVGV	621
Qy	489	GYDRTQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVAWMGNSGGGGTTPRAVPGRVVALGOT	548
Db	622	GYDRTQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVAWMGNSGGGGTTPRAVPGRVVALGOT	681
Qy	549	VQASDSLTCGAETLNGLIQFDAAIQPDGSGPVVNGLGQVWGMNTAAS	596
Db	682	VQASDSLTCGAETLNGLIQFDAAIQPDGSGPVVNGLGQVWGMNTAAS	729

```

RESULT 12
US-10-369-983-17
/ Sequence 17, Application US/10369983
/ Publication No. US20030235593A1
/ GENERAL INFORMATION:
/ APPLICANT: Skeiky, Yasir
/ APPLICANT: Guderian, Jeff
/ APPLICANT: Reed, Steven
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
/ FILE REFERENCE: 014058-009081US
/ CURRENT APPLICATION NUMBER: US/10/369,983
/ CURRENT FILING DATE: 2003-02-18
/ PRIOR APPLICATION NUMBER: US 60/357,351
/ PRIOR FILING DATE: 2002-02-15
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: PatentIn ver. 2.1
/ SEQ ID NO 17
/ LENGTH: 1022
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: fusion protein
/ -OTHER INFORMATION: MTB102tm2F (MTB102PTM2, MTB72F-hTCC#1)
US-10-369-993-17

```

Qy	69	SSAGLWAAASPYYVMWSVTAGQELTAQVVRVAAAAVETAYGLTVPPPVTAENRAELMI	128
Db	202	SSAGLWAAASPYYVMWSVTAGQELTAQVVRVAAAAVETAYGLTVPPPVTAENRAELMI	261
Qy	129	LIATNLLGONTPAI AVNEABYGENWAOQAAAFVGYAAATATATATLLPFEEAPEMTSAGG	188
Db	262	LIATNLLGONTPAI AVNEABYGENWAOQAAAFVGYAAATATATATLLPFEEAPEMTSAGG	321
Qy	189	LLFQAAAVEASDTAAANQLMNNVPQALQOQLOPTQGTTPSSKLGGLWKTVPSPRSPISN	248
Db	322	LLFQAAAVEASDTAAANQLMNNVPQALQOQLOPTQGTTPSSKLGGLWKTVPSPRSPISN	381
Qy	249	MVSMANNHSMTNSGYSMTNTLS SMLKGFAPAAAAQAVQTAAQNGVRAMSLGSSLSGG	308
Db	382	MVSMANNHSMTNSGYSMTNTLS SMLKGFAPAAAAQAVQTAAQNGVRAMSLGSSLSGG	441
Qy	309	LGGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV	368
Db	442	LGGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV	501
Qy	369	GQNGARAGGSLGVLRVPRPYVMHSPAAAGDIAPPALSDRRPADFPALPLDPSAMVAQV	428
Db	502	GQNGARAGGSLGVLRVPRPYVMHSPAAAGDIAPPALSDRRPADFPALPLDPSAMVAQV	561
Qy	429	GPQVVMINTKLGYNNVAGAGTGVIDPNGVVLNNHVIAGATDINAFSVGSGQTYGVDDV	488
Db	562	GPQVVMINTKLGYNNVAGAGTGVIDPNGVVLNNHVIAGATDINAFSVGSGQTYGVDDV	621
Qy	489	GYDRTQDVAVLQURGAGGLPSAAIIGGSVA VGEPPVAMNSGGGGTTPRVPGRVVAJLQGT	548
Db	622	GYDRTQDVAVLQURGAGGLPSAAIIGGSVA VGEPPVAMNSGGGGTTPRVPGRVVAJLQGT	681
Qy	549	VQASDSLTA GAEETLNGLI OFDAAIQPCDGGPPVNGLGQTVGMNTAAS	596
Db	682	VQASDSLTA GAEETLNGLI OFDAAIQPCDGGPPVNGLGQTVGMNTAAS	729

```

RESULT 13
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-983-16

```

	Query Match	98.0%;	Score 2946;	DB 15;	Length 1154;
	Best Local Similarity	100.0%;	Pred. NO. 2.2e-197;		
	Matches 588;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	9	MYDFGALPPEINSARMYAPGSGASIVAAQMWDSVASDLFSAASAPQSVVWGLTVGSWIG	68		
Db	142	MYDFGALPPEINSARMYAPGSGASIVAAQMWDSVASDLFSAASAPQSVVWGLTVGSWIG	201		
QY	69	SSAGLMLVAAASPYVAMWSVTAGCAELTAAOVVRAAAAYETAYGLTVPPVPVIAENRAELMI	128		

202 SSAGLWVAASPPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
 129 LIATNLLGQNTPAIAVNEAYGEMWAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 188
 262 LIATNLLGQNTPAIAVNEAYGEMWAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 321
 189 LLEQAAAVEASDTAAANQMLMNVPAALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 248
 322 LLEQAAAVEASDTAAANQMLMNVPAALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 381
 249 MVSMAHNMHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
 382 MVSMAHNMHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
 309 LGGGVAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 368
 442 LGGGVAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 501
 369 GQMGARAGGGLSGVLRVPPPYVMPHSPAAQDIAPPALSOQDRFADFPALPLDPSAMVAQV 428
 502 GQMGARAGGGLSGVLRVPPPYVMPHSPAAQDIAPPALSOQDRFADFPALPLDPSAMVAQV 561
 429 GPQVNNINTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVV 488
 562 GPQVNNINTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVV 621
 489 GYDRTQDVAVLQLRGAGLPSAATGGGVAVGEPPVAMGNSGGGQTPRAVPGRVVALGQT 548
 622 GYDRTQDVAVLQLRGAGLPSAATGGGVAVGEPPVAMGNSGGGQTPRAVPGRVVALGQT 581
 549 VQASDSLTGAETLNGLIQFDAAIQPCDGGPVGNGLGQVVGWMTAAS 596
 682 VQASDSLTGAETLNGLIQFDAAIQPCDGGPVGNGLGQVVGWMTAAS 729

RESULT 14
 US-09-886-349A-18
 ; Sequence 18, Application US/09886349A
 ; Publication No. US20040086523A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009070US
 ; CURRENT APPLICATION NUMBER: US/09/886,349A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: US 09/597,796
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: US 60/265,737
 ; PRIOR FILING DATE: 2001-02-01
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 18
 ; TYPE: PRT
 ; LENGTH: 729
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
 ; OTHER INFORMATION: (Ra12-TbHp-Ra35MutSA)
 US-09-886-349A-18

Query Match 97.9%; Score 2943; DB 12; Length 729;
 Best Local Similarity 99.8%; Pred. No. 2e-197;
 Matches 587; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 9 MVDGALPPEINSMYAGPAGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGWSWIG 68
 142 MVDGALPPEINSMYAGPAGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGWSWIG 201
 69 SSAGLWVAASPPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128

202 SSAGLWVAASPPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
 129 LIATNLLGQNTPAIAVNEAYGEMWAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 188
 262 LIATNLLGQNTPAIAVNEAYGEMWAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 321
 189 LLEQAAAVEASDTAAANQMLMNVPAALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 248
 322 LLEQAAAVEASDTAAANQMLMNVPAALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 381
 249 MVSMAHNMHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
 382 MVSMAHNMHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
 309 LGGGVAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 368
 442 LGGGVAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 501
 369 GQMGARAGGGLSGVLRVPPPYVMPHSPAAQDIAPPALSOQDRFADFPALPLDPSAMVAQV 428
 502 GQMGARAGGGLSGVLRVPPPYVMPHSPAAQDIAPPALSOQDRFADFPALPLDPSAMVAQV 561
 429 GPQVNNINTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVV 488
 562 GPQVNNINTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVVV 621
 489 GYDRTQDVAVLQLRGAGLPSAATGGGVAVGEPPVAMGNSGGGQTPRAVPGRVVALGQT 548
 622 GYDRTQDVAVLQLRGAGLPSAATGGGVAVGEPPVAMGNSGGGQTPRAVPGRVVALGQT 581
 549 VQASDSLTGAETLNGLIQFDAAIQPCDGGPVGNGLGQVVGWMTAAS 596
 682 VQASDSLTGAETLNGLIQFDAAIQPCDGGPVGNGLGQVVGWMTAAS 729

RESULT 15
 US-10-098-732A-18
 ; Sequence 18, Application US/10098732A
 ; Publication No. US20030175294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Brannon, Mark
 ; APPLICANT: Guderian, Jeffrey
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; TITLE OF INVENTION: Leishmania Antigen
 ; FILE REFERENCE: 014058-012010US
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: US 60/275,837
 ; PRIOR FILING DATE: 2001-03-13
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 18
 ; TYPE: PRT
 ; LENGTH: 729
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
 ; OTHER INFORMATION: (Ra12-TbH9-Ra35MutSA)
 US-10-098-732A-18

Query Match 97.9%; Score 2943; DB 14; Length 729;
 Best Local Similarity 99.8%; Pred. No. 2e-197;
 Matches 587; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 9 MVDGALPPEINSMYAGPAGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGWSWIG 68
 142 MVDGALPPEINSMYAGPAGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGWSWIG 201
 69 SSAGLWVAASPPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128

```
Db      202  SSAGLMVAASPVTAMSVTAGAEELTAAQVRVAAAYETAYGLTVPPPEVIAENRAELMI 261
QY      129  LIATNLLGQNTPAIAVNEAEYGEEMWAQDAAMFGYAAATATATATILPPEEAPEMTSAGG 188
Db      262  LIATNLLGQNTPAIAVNEAEYGEEMWAQDAAMFGYAAATATATATILPPEEAPEMTSAGG 321
QY      189  LLEQAAVEEASDSTAANQIMNNVPOALQOLAQPTQGITPSSKLGGLWKTVSPHRSPISN 248
Db      322  LLEQAAVEEASDSTAANQIMNNVPOALQOLAQPTQGITPSSKLGGLWKTVSPHRSPISN 381
QY      249  MYSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 308
Db      382  MYSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAPAAAQAVOTAAQNGVRAMSSLGSSLGSSG 441
QY      309  LGGGVAANLGRAASVGSLSVPOAWAAANOAVTPAARALPLTSLTSAERGPQOMLGGLPV 368
Db      442  LGGGVAANLGRAASVGSLSVPOAWAAANOAVTPAARALPLTSLTSAERGPQOMLGGLPV 501
QY      369  GQMGARAGGGLSGVLRVPRPYMPHSPAAAGDIAPPALSCDFADFPALPLDPSAMVAQV 428
Db      502  GQMGARAGGGLSGVLRVPRPYMPHSPAAAGDIAPPALSCDFADFPALPLDPSAMVAQV 561
QY      429  GPQVNVNINTKLYNNNAVAGTGVIDPNGVILTNHVIAGATDINAFVSGSQTYGVDVV 488
Db      562  GPQVNVNINTKLYNNNAVAGTGVIDPNGVILTNHVIAGATDINAFVSGSQTYGVDVV 621
QY      489  GYDRTQDVAVLOIRGAGGLPSAIIGGGVAVGEPPVAMGNSGGQGTTPRAVPCRVALCQT 548
Db      622  GYDRTQDVAVLOIRGAGGLPSAIIGGGVAVGEPPVAMGNSGGQGTTPRAVPCRVALCQT 681
QY      549  VQASDSLTAEEETLNLGLIQFDAAIQPGSDGPPVNVNGLGOVVGMTAAS 596
Db      682  VQASDSLTAEEETLNLGLIQFDAAIQPGDAGPPVNVNGLGOVVGMTAAS 729
```

Search completed: June 30, 2004, 17:14:46
Job time : 48.4747 secs

1. The first part of the document is a list of names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The list includes names such as Mr. John A. Smith, Mr. James B. Jones, and Mr. Robert C. Brown. The addresses are listed in the same order as the names.

2. The second part of the document is a list of the names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The list includes names such as Mr. John A. Smith, Mr. James B. Jones, and Mr. Robert C. Brown. The addresses are listed in the same order as the names.

3. The third part of the document is a list of the names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The list includes names such as Mr. John A. Smith, Mr. James B. Jones, and Mr. Robert C. Brown. The addresses are listed in the same order as the names.

4. The fourth part of the document is a list of the names and addresses of the members of the committee. The names are listed in alphabetical order, and the addresses are listed below each name. The list includes names such as Mr. John A. Smith, Mr. James B. Jones, and Mr. Robert C. Brown. The addresses are listed in the same order as the names.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 18.1302 Seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796C-10
Perfect score: 3007
Sequence: 1 HMEHHHHMVDFFGALPPEIN.....SGGPVNVGLQGVVGMNTAAS 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.psp.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.psp.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.psp.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.psp.*
5: /cgn2_6/ptodata/2/iaa/PCRTUS_COMB.psp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.psp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3007	100.0	596	4	US-09-287-849-26
2	2941	97.8	729	4	US-09-223-040-2
3	2941	97.8	729	4	US-09-287-849-2
4	2033.5	67.6	600	4	US-09-287-849-22
5	1949	64.8	391	3	US-08-818-112-107
6	1949	64.8	391	4	US-08-818-111-102
7	1949	64.8	391	4	US-09-056-556-107
8	1949	64.8	391	4	US-09-072-967-102
9	1949	64.8	391	4	US-09-072-967-107
10	1652.5	55.0	396	3	US-08-818-112-111
11	1652.5	55.0	396	4	US-08-818-111-106
12	1652.5	55.0	396	4	US-09-056-556-111
13	1652.5	55.0	396	4	US-09-072-956-106
14	1652.5	55.0	396	4	US-09-072-967-111
15	1486.5	49.4	359	3	US-08-818-112-109
16	1486.5	49.4	359	4	US-08-818-111-104
17	1486.5	49.4	359	4	US-09-056-556-109
18	1486.5	49.4	359	4	US-09-072-956-104
19	1486.5	49.4	358	4	US-09-072-967-109
20	1189	39.5	358	4	US-09-287-849-8
21	1187	39.5	263	3	US-08-818-112-91
22	1187	39.5	263	4	US-08-818-111-92
23	1187	39.5	263	4	US-09-056-556-91
24	1187	39.5	263	4	US-09-072-956-92
25	1187	39.5	263	4	US-09-072-967-91
26	983	32.7	355	3	US-08-818-112-79
27	983	32.7	355	4	US-08-818-111-80

28	983	32.7	355	4	US-09-056-556-79	Sequence 79, Appl
29	983	32.7	355	4	US-09-072-956-80	Sequence 80, Appl
30	983	32.7	355	4	US-09-072-967-79	Sequence 79, Appl
31	770.5	25.6	400	4	US-09-073-009-126	Sequence 126, App
32	770.5	25.6	400	4	US-09-073-010-126	Sequence 126, App
33	607.5	20.2	710	4	US-09-287-849-16	Sequence 16, Appl
34	607.5	20.2	856	4	US-09-287-849-12	Sequence 12, Appl
35	606.5	20.2	423	4	US-09-073-009-142	Sequence 142, App
36	606.5	20.2	423	4	US-09-073-010-142	Sequence 142, App
37	450.5	15.0	943	4	US-09-477-135A-131	Sequence 131, App
38	424	14.1	141	4	US-09-073-009-15	Sequence 15, Appl
39	424	14.1	141	4	US-09-073-010-15	Sequence 15, Appl
40	381.5	12.7	204	4	US-08-311-731A-57	Sequence 57, Appl
41	377.5	12.6	208	4	US-08-311-731A-208	Sequence 208, App
42	328.5	10.9	371	4	US-09-050-739-92	Sequence 92, Appl
43	325.5	10.8	368	3	US-08-818-112-114	Sequence 114, App
44	325.5	10.8	368	4	US-08-818-111-109	Sequence 109, App
45	325.5	10.8	368	4	US-09-056-556-114	Sequence 114, App

ALIGNMENTS

RESULT 1
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillot, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match		100.0%;	Score 3007;	DB 4;	Length 596;
Best Local Similarity		100.0%;	Pred. No. 1.3e-219;	Indels 0;	Gaps 0;
Matches 596;		Conservative 0;	Mismatches 0;		
Qy	1	HMEHHHHMVDFFGALPPEINARMYAGPGSASISIAAAQWMDSVASDLFGAASAFQSVWVG	60		
Db	1	HMEHHHHMVDFFGALPPEINARMYAGPGSASISIAAAQWMDSVASDLFGAASAFQSVWVG	60		
Qy	61	LTGSGWIGSAGLMVAASPYVAMSVTAQBELTAQVRVAAAAYETAYGLTVPPEVIA	120		
Db	61	LTGSGWIGSAGLMVAASPYVAMSVTAQBELTAQVRVAAAAYETAYGLTVPPEVIA	120		
Qy	121	ENRAELMILATNLGONTPALAVNEABYCEMWAQDAAMFGVAAATATATATLLPPEEA	180		

Db 121 ENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEA 180
Qy 181 PENTSAAGLLLEQAAAEVEASDTAAANQOLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTYS 240
Db 181 PENTSAAGLLLEQAAAEVEASDTAAANQOLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTYS 240
Qy 241 PHSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSL 300
Db 241 PHSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSL 300
Qy 301 GSSLGSSGLCGGVAANLGRAASVGSLSVPOAWAAANQAVTPPAARALPLTSLTSAEERGGP 360
Db 301 GSSLGSSGLCGGVAANLGRAASVGSLSVPOAWAAANQAVTPPAARALPLTSLTSAEERGGP 360
Qy 361 QMLGGLPVGQMGARAGGLSGVLRVPPRPVYMPHSPAAAGDIAPPALSDQRFADFPALPLD 420
Db 361 QMLGGLPVGQMGARAGGLSGVLRVPPRPVYMPHSPAAAGDIAPPALSDQRFADFPALPLD 420
Qy 421 PSAMVAQGVQVNNINTKLYNNNAVAGAGTGIVIDPENGVLVLTNNHVIAGATDINAFSVGSG 480
Db 421 PSAMVAQGVQVNNINTKLYNNNAVAGAGTGIVIDPENGVLVLTNNHVIAGATDINAFSVGSG 480
Qy 481 QTVGVVDVDRDQDVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGCGGTPRAVPG 540
Db 481 QTVGVVDVDRDQDVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGCGGTPRAVPG 540
Qy 541 RVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAAS 596
Db 541 RVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAAS 596

RESULT 2
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6543522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2

Query Match 97.8%; Score 2941; DB 4; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.7e-214; Indels 0; Gaps 0;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSWG 68
Db 142 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSWG 201

Qy 69 SSAGLMVAASPPVYVWMSVTAGQELTAACQVRVAAAAYETAYGLTVPPIAENRAELMI 128
Db 202 SSAGLMVAASPPVYVWMSVTAGQELTAACQVRVAAAAYETAYGLTVPPIAENRAELMI 261

Qy 129 LIATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAG 188
Db 262 LIATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAG 321

Qy 189 LLEQAAAEVEASDTAAANQOLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTYSPPHRSPI 248

Db 322 LLEQAAAEVEASDTAAANQOLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTYSPPHRSPI 381
Qy 249 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSG 308
Db 382 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSG 441
Qy 309 LGGVAAANLGRAASVGSLSVPOAWAAANQAVTPPAARALPLTSLTSAEERGGPQMLGSLPV 368
Db 442 LGGVAAANLGRAASVGSLSVPOAWAAANQAVTPPAARALPLTSLTSAEERGGPQMLGSLPV 501
Qy 369 GQMGARAGGLSGVLRVPPRPVYMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGLSGVLRVPPRPVYMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
Qy 429 GPQVNNINTKLYNNNAVAGAGTGIVIDPENGVLVLTNNHVIAGATDINAFSVGSGQTYGVDDV 488
Db 562 GPQVNNINTKLYNNNAVAGAGTGIVIDPENGVLVLTNNHVIAGATDINAFSVGSGQTYGVDDV 621
Qy 489 GYDRTQDVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGCGGTPRAVPGRVWALGOT 548
Db 622 GYDRTQDVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGCGGTPRAVPGRVWALGOT 681
Qy 549 VOASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAAS 596
Db 682 VOASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAAS 729

RESULT 3
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6827198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-2

Query Match 97.8%; Score 2941; DB 4; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.7e-214; Indels 0; Gaps 0;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSWG 68
Db 142 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSWG 201

Qy 69 SSAGLMVAASPPVYVWMSVTAGQELTAACQVRVAAAAYETAYGLTVPPIAENRAELMI 128

Db 202 SSAGLWVAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 261
Qy 129 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPFEAEPMTSAGG 188
Db 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAEPMTSAGG 321
Qy 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQCTTSSKLGLLWKTVPSPHSPISN 248
Db 322 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQCTTSSKLGLLWKTVPSPHSPISN 381
Qy 249 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGSSG 308
Db 382 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGSSG 441
Qy 309 LGGVAAANLGRAASVGLSVPOQAAANQAVTPAARALPLTSLTSAERPGQMLGLPV 368
Db 442 LGGVAAANLGRAASVGLSVPOQAAANQAVTPAARALPLTSLTSAERPGQMLGLPV 501
Qy 369 GOMCARAGGLSGVLRVPPRPVYVPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAV 428
Db 502 GOMCARAGGLSGVLRVPPRPVYVPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAV 561
Qy 429 GPQVNNITKLYNNVAGAGTGVIDPENGVLNNHVIAGATDINAFSVSGSGTGYGVVV 488
Db 562 GPQVNNITKLYNNVAGAGTGVIDPENGVLNNHVIAGATDINAFSVSGSGTGYGVVV 621
Qy 489 GYDRTQDAVQLRGAGLSPAAIGGVAQVGEVPMVAMNSGGGGTTPRAVGRVVALGOT 548
Db 622 GYDRTQDAVQLRGAGLSPAAIGGVAQVGEVPMVAMNSGGGGTTPRAVGRVVALGOT 681
Qy 549 VOASDLSGTAEETLNGLIQDAIOPGDSGGPVVNGLOVVGMMNTAAS 596
Db 682 VOASDLSGTAEETLNGLIQDAIOPGDSGGPVVNGLOVVGMMNTAAS 729

RESULT 4

US-09-287-849-22
; Sequence 22 Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 57.6%; Score 2033.5; DB 4; Length 600;
Best Local Similarity 72.5%; Pred. No. 6,7e-146;

Matches 440; Conservative 19; Mismatches 81; Indels 67; Gaps 10;
Qy 1 HHHHHHHHVVDFGALPPPEINSARMYAGCSASLVAAAQWDSVASDLPSAASAFOSVVMG 60
Db 1 HHHHHHHHVVDFGALPPPEINSARMYAGCSASLVAAAQWDSVASDLPSAASAFOSVVMG 60
Qy 61 LTVGSWIGSSAGLWVAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIA 120
Db 61 LTVGSWIGSSAGLWVAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIA 120
Qy 121 ENRAELMLIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLPPFEA 180
Db 121 ENRAELMLIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLPPFEA 180
Qy 181 PMTSAGLLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQCTTSSKLGLLWKTVS 240
Db 181 PMTSAGLLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQCTTSSKLGLLWKTVS 240
Qy 241 PHRSPI SNMVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSL 300
Db 241 PHRSPI SNMVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSL 300
Qy 301 GSSLGSSGLGGVAAANLGRAASVGLSVPOQAAANQAVTPAARALPLTSLTSAERPG 360
Db 301 GSSLGSSGLGGVAAANLGRAASVGLSVPOQAAANQAVTPAARALPLTSLTSAERPG 360
Qy 361 QMLGLPLVQMGARAGGLSGVLRVPPRPVYVPHSPAAAGDIAP----- 403
Db 361 QMLGLPLVQMGARAGGLSGVLRVPPRPVYVPHSPAAAGKLDPPDAVINTTCNYQVVA 420
Qy 404 -PALSDQRFADFPALPLDPS-----AMVAQVGPVNNITKLYNNVAGAG 448
Db 421 LNATDPGAAQFNAPSVQSYLRNFLLAAPPQRAAQAQL-QAVPGAAQYIGLVESVAGS 479
Qy 449 TGVIDPENGVLNNHVIAGATDINAFSVSGSGTGYGVVVNGLOVVGMMNTAAS 508
Db 480 CN-----NYELMTINYQFG---DVDA---HGMIRAQAASLEAEHQAIVRDLAAGDFW 527
Qy 509 SAAICGGVAVGEPVVMG-----NSGGGGTTPRAVGRVVALGOTVOASDLSLGA 558
Db 528 GGA--GSVACQEFITQLGRNFQVIYEQANAHGQ-----KVOAGNNNAQTDTSVAGS 576
Qy 559 E-ETLING 564
Db 577 SWATNG 583

RESULT 5

US-08-818-112-107
; Sequence 107 Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/08/818,112
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn Ver. 2.1
; ADDRESS: SEED and BERY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; MEDIUM TYPE: IBM PC compatible
 ; FILING DATE: 13-MAR-1997
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 107:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-818-112-107

Query Match 64.8%; Score 1949; DB 3; Length 391;
 Best Local Similarity 100.0%; Pred. No. 9.4e-140;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
 DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 QY 69 SSAGLWVAASPYVWAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 DB 61 SSAGLWVAASPYVWAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 129 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPEMTSAGG 188
 DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPEMTSAGG 180
 QY 189 LLEQAAAVEASDTAAANQNMNVPAALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
 DB 181 LLEQAAAVEASDTAAANQNMNVPAALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
 QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSLGSSG 308
 DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSLGSSG 300
 QY 309 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 368
 DB 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 360
 QY 369 GQMGARAGGGLSGVLRVPPRPYVWPHSPAAG 399
 DB 361 GQMGARAGGGLSGVLRVPPRPYVWPHSPAAG 391

RESULT 6
 US-08-818-111-102
 ; Sequence 102, Application US/08818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA

; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: 13-MAR-1997
 ; APPLICATION NUMBER: US/08/818,111
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 102:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-818-111-102

Query Match 64.8%; Score 1949; DB 4; Length 391;
 Best Local Similarity 100.0%; Pred. No. 9.4e-140;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
 DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 QY 69 SSAGLWVAASPYVWAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 DB 61 SSAGLWVAASPYVWAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 129 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPEMTSAGG 188
 DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPEMTSAGG 180
 QY 189 LLEQAAAVEASDTAAANQNMNVPAALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
 DB 181 LLEQAAAVEASDTAAANQNMNVPAALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
 QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSLGSSG 308
 DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSLGSSG 300
 QY 309 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 368
 DB 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 360
 QY 369 GQMGARAGGGLSGVLRVPPRPYVWPHSPAAG 399
 DB 361 GQMGARAGGGLSGVLRVPPRPYVWPHSPAAG 391

RESULT 7
 US-09-056-556-107
 ; Sequence 107, Application US/09056556
 ; Patent No. 6350456
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 ; NUMBER OF SEQUENCES: 241
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington

```

; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-107

```

```

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
DB 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 69 SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
DB 61 SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 129 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 188
DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180

QY 189 LLEQAAVVEASDTAAANQLMNVPOALQOLQAQPTGTTSSKLGGLWKTVPSPHRSPI 248
DB 181 LLEQAAVVEASDTAAANQLMNVPOALQOLQAQPTGTTSSKLGGLWKTVPSPHRSPI 240

QY 249 MYSMANNHMTNSGVSMTNTLSSMLKGPAPAAAQAVTAAQNGVRAAMSSLSGSGG 308
DB 241 MYSMANNHMTNSGVSMTNTLSSMLKGPAPAAAQAVTAAQNGVRAAMSSLSGSGG 300

QY 309 LGGVVAANLGRAASVGSLSVPOQAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPV 368
DB 301 LGGVVAANLGRAASVGSLSVPOQAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPV 360

```

RESULT 8

```

US-09-072-596-102
; Sequence 102, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.

```

```

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-102

```

```

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
DB 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 69 SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
DB 61 SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 129 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 188
DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180

QY 189 LLEQAAVVEASDTAAANQLMNVPOALQOLQAQPTGTTSSKLGGLWKTVPSPHRSPI 248
DB 181 LLEQAAVVEASDTAAANQLMNVPOALQOLQAQPTGTTSSKLGGLWKTVPSPHRSPI 240

QY 249 MYSMANNHMTNSGVSMTNTLSSMLKGPAPAAAQAVTAAQNGVRAAMSSLSGSGG 308
DB 241 MYSMANNHMTNSGVSMTNTLSSMLKGPAPAAAQAVTAAQNGVRAAMSSLSGSGG 300

QY 309 LGGVVAANLGRAASVGSLSVPOQAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPV 368
DB 301 LGGVVAANLGRAASVGSLSVPOQAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPV 360

```

RESULT 9

```

US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.

```

APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072.967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-107

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
DB 1 MVDGALPPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLWAAASPYVWMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 128
DB 61 SSAGLWAAASPYVWMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 120
QY 129 LIATNLLGONTPAI AVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188
DB 121 LIATNLLGONTPAI AVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
QY 189 LLEQAAVEEASDTAAANQLMNNVFPALQQAQPTQGTTPSSKLGGLWKTVPVHRSPISN 248
DB 181 LLEQAAVEEASDTAAANQLMNNVFPALQQAQPTQGTTPSSKLGGLWKTVPVHRSPISN 240
QY 249 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSG 308
DB 241 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLGSSG 300
QY 309 LGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARPGQMLGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARPGQMLGLPV 360
QY 369 GQMGARAGGGLSGVLRVPPRPMVPHSPAAG 399
DB 361 GQMGARAGGGLSGVLRVPPRPMVPHSPAAG 391

RESULT 10
US-08-818-112-111
Sequence 111, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818.112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
US-08-818-112-111

Query Match 55.0%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 9 MVDGALPPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
DB 1 MVDGALPPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLWAAASPYVWMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 128
DB 61 SSAGLWAAASPYVWMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 120
QY 129 LIATNLLGONTPAI AVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188
DB 121 LIATNLLGONTPAI AVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
QY 189 LLEQAAVEEASDTAAANQLMNNVFPALQQAQPTQGTTPSSKLGGLWKTVPVHRSPISN 248
DB 181 LLEQAAVEEADTAAANQLMNNVFPALQQAQPTKSIWFPDQSELWKALSPHLSPLSN 240
QY 249 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSS----LGSSL 304
DB 241 IVSMLNHNVTNSGVSMNTLHSMKGFAP-AAAQAVETAQNGVQAMSSLGSSG 299
QY 305 GSSGLGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARPGQMLG 364

Db 300 GSSGLGAGVAVANLGRAASVGSLSVPQAWAANAQVTPAARALFLTLTSAQAOTAPGHMLG 359
QY 365 GLPVGQMGARAG--GGLSGVLRVPPRPVYMPHSPAAG 399
Db 360 GLPLQLTNSGGGFGVSNALRMPPRAYVMPVPAAG 396
RESULT 11
US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-111-106
Query Match 55.0%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 9 MVDGALPPEINSARMYAGFGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTGVSWIG 68
Db 1 VVDGALPPEINSARMYAGFGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTGVSWIG 60
QY 69 SSAGLMVAASPPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128
Db 61 SSAGLMVAASPPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 129 LIATNLLGONTPTAIVNEAYEGEMWQADAAAFGYAAATATATATILPPEAPMTSAGG 188
Db 121 LIATNLLGONTPTAIVNEAYEGEMWQADAAAFGYAAATATATATILPPEAPMTSAGG 180
QY 189 LLEQAAVVEASDTAAANQLMNNVPQALQQLAQTGTTPTSSKLGGLWKTVPSPHRSPISN 248
Db 181 LLEQAVAVEAIDTAAANQLMNNVPQALQQLAQTGTTPTSSKLGGLWKTVPSPHRSPISN 240
QY 249 MYSMANNHVMNTNSGVSMNTLSSMLKGFAPAAQAQVTAQAQNGVRMSS-----LGSSSL 304

Db 241 IVSMLNHVMNTNSGVSMNTLSSMLKGFAP-AAQAQVETAQAQNGVQAQMSILGSSSL 299
QY 305 GSSGLGAGVAVANLGRAASVGSLSVPQAWAANAQVTPAARALFLTLTSAQAOTAPGHMLG 364
Db 300 GSSGLGAGVAVANLGRAASVGSLSVPQAWAANAQVTPAARALFLTLTSAQAOTAPGHMLG 359
QY 365 GLPVGQMGARAG--GGLSGVLRVPPRPVYMPHSPAAG 399
Db 360 GLPLQLTNSGGGFGVSNALRMPPRAYVMPVPAAG 396
RESULT 12
US-09-056-556-111
; Sequence 111, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-111
Query Match 55.0%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 9 MVDGALPPEINSARMYAGFGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTGVSWIG 68
Db 1 VVDGALPPEINSARMYAGFGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTGVSWIG 60
QY 69 SSAGLMVAASPPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128
Db 61 SSAGLMVAASPPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 129 LIATNLLGONTPTAIVNEAYEGEMWQADAAAFGYAAATATATATILPPEAPMTSAGG 188
Db 121 LIATNLLGONTPTAIVNEAYEGEMWQADAAAFGYAAATATATATILPPEAPMTSAGG 180
QY 189 LLEQAAVVEASDTAAANQLMNNVPQALQQLAQTGTTPTSSKLGGLWKTVPSPHRSPISN 248
Db 181 LLEQAVAVEAIDTAAANQLMNNVPQALQQLAQTGTTPTSSKLGGLWKTVPSPHRSPISN 240

TREATME

QY 249 MVSNNHMTNMGVMTNTLSSMLKGFAPAAAQAQVOTAAQNGVAMSS-----LGSSL 304
Db 241 IVSMLNHNHVTNMGVSMSTLHSLKGFAP-AAAQAVETAQNGVQVAMSSLSQSLGSSL 299
QY 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGOMLG 364
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAQAPGHMLG 359
QY 365 GLPVQMGARAG--CGLSGLVLRVPPRYVMPHSPAAG 399
Db 360 GLPLGQLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396
RESULT 13
US-09-072-596-106
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-596-106
Query Match 55.0%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 9 MVDGALPPEINARMYAGPGSASLVAAAQWDSVADLSFAAFAQSVVWGLTVGSMIG 68
Db 1 VVDFGALPPEINARMYAGPGSASLVAAAQWDSVADLSFAAFAQSVVWGLTVGSMIG 60
QY 69 SSAGLWVAAPGYVWMSVTAGQAEELTAQVRAAAAAYETAYGLTVPPVIAENRAELMI 128
Db 61 SSAGLWVAAPGYVWMSVTAGQAEELTAQVRAAAAAYETAYGLTVPPVIAENRAELMI 120
QY 129 LIATNLLQNTPAIAVNEAEYGEWMAQDAAMAFGAAATATATATLTLPPFEAPENTAGG 188

Db 121 LIATNLLQNTPAIAVNEAEYGEWMAQDAAMAFGAAATATATATLTLPPFEAPLITNPGG 180
QY 189 LLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTOGTTSSKLGGLWKTVPSPHRSPLSN 248
Db 181 LLEQAAVEEADIDTAANQLMNNVPOALQOLAQPTKSIWFPDQJSELWKAISPHLSPLSN 240
QY 249 MVSNNHMTNMGVMTNLTSSMLKGFAPAAAQAQVOTAAQNGVAMSS-----LGSSL 304
Db 241 IVSMLNHNHVTNMGVSMSTLHSLKGFAP-AAAQAVETAQNGVQVAMSSLSQSLGSSL 299
QY 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGOMLG 364
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAQAPGHMLG 359
QY 365 GLPVQMGARAG--CGLSGLVLRVPPRYVMPHSPAAG 399
Db 360 GLPLGQLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396
RESULT 14
US-09-072-967-111
; Sequence 111, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-967-111
Query Match 55.0%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 9 MVDGALPPEINARMYAGPGSASLVAAAQWDSVADLSFAAFAQSVVWGLTVGSMIG 68

```
Db 1 VVDFGALPPEINSAARMYAGPGSASLVAAAKQWDSVASDLFSAASAFQSVVWGLTIGSWG 60
Qy 69 SSAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
Db 61 SSAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
Qy 129 LIATNLLGQNTPAIAVNEAEYGENMAQDAAMFGYAAATATATATATATATATATATATAT 188
Db 121 LIATNLLGQNTPAIAVNEAEYGENMAQDAAMFGYAAATATATATATATATATATATATAT 180
Qy 189 LLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGTTSSKLGGLWKTVPVPHRSPISN 248
Db 181 LLEQAAVEEADITAAANQLMNNVPOALQOLAQPTGTTSSKLGGLWKTVPVPHRSPISN 240
Qy 249 MVSANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSS----LGSSL 304
Db 241 IVSMLNHNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSS----LGSSL 299
Qy 305 GSSGLGGVAAANLGRAASVGSLSVPOAAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 364
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 359
Qy 365 GLPVGQMGARAG--GGLSGVLVPPVPPVYVPHSPFAAG 399
Db 360 GLPLGQLTNSGGGFGGVSNALRMPPRAYVMPVPAAG 396
```

RESULT 15

```
US-08-818-112-109
; Sequence 109, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-112-109
```

```
Query Match 49.4%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 8.8e-105;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;
Qy 9 VVDFGALPPEINSAARMYAGPGSASLVAAAKQWDSVASDLFSAASAFQSVVWGLTIGSWG 68
Db 1 VVDFGALPPEINSAARMYAGPGSASLVAAAKQWDSVASDLFSAASAFQSVVWGLTIGSWG 60
Qy 69 SSAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
Db 61 SSAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
Qy 129 LIATNLLGQNTPAIAVNEAEYGENMAQDAAMFGYAAATATATATATATATATATATATAT 188
Db 121 LIATNLLGQNTPAIAVNEAEYGENMAQDAAMFGYAAATATATATATATATATATATATAT 180
Qy 189 LLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGTTSSKLGGLWKTVPVPHRSPISN 248
Db 181 LLEQAAVEEADITAAANQLMNNVPOALQOLAQPTGTTSSKLGGLWKTVPVPHRSPISN 240
Qy 249 MVSANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSS----LGSSL 304
Db 241 VSSANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSS----LGSSL 299
Qy 305 GSSGLGGVAAANLGRAASVGSLSVPOAAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 364
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 359
```

Search completed: June 30, 2004, 16:55:20
Job time : 19.1302 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 61.2853 Seconds
(without alignments)
2747.774 Million cell updates/sec

Title: US-09-597-796C-10

Perfect score: 3007

Sequence: 1 RMHHHHHVMDFGALPPEIN.....SGGPVNLGLGVGVGMNTAAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3007	100.0	596	2 AAY32070	Aay32070 Mycobacte
2	3007	100.0	596	5 AAE29710	Aae29710 Mycobacte
3	3007	100.0	596	5 AAE17574	Aae17574 Mycobacte
4	3007	100.0	599	5 AAU74599	AAU74599 Antigenic
5	2946	98.0	729	4 AAO22142	AAO22142 Ra12-H9-3
6	2946	98.0	813	7 ADA26367	ADA26367 Mycobacte
7	2946	98.0	825	7 ADA26366	ADA26366 Mycobacte
8	2946	98.0	875	7 ADA26365	ADA26365 Mycobacte
9	2946	98.0	930	5 AAE29731	Aae29731 Mycobacte
10	2946	98.0	930	7 ADA26364	ADA26364 Mycobacte
11	2946	98.0	1016	7 ADA26370	ADA26370 M. bovis
12	2946	98.0	1022	7 ADA26369	ADA26369 Mycobacte
13	2946	98.0	1154	7 ADA26368	ADA26368 Mycobacte
14	2943	97.9	729	5 AAE29709	Aae29709 Mycobacte
15	2943	97.9	729	5 AAE17573	Aae17573 Mycobacte
16	2943	97.9	729	5 ADA26374	ADA26374 Mycobacte
17	2941	97.8	729	5 AAE29708	Aae29708 Mycobacte
18	2941	97.8	729	5 AAE17572	Aae17572 Mycobacte
19	2936	97.6	729	7 ADA26373	ADA26373 Mycobacte
20	2923	97.2	729	2 AAY32059	Aay32059 Mycobacte
21	2894.5	96.3	726	5 AAU74588	AAU74588 Antigenic
22	2033.5	67.6	600	2 AAY32068	Aay32068 Mycobacte
23	2033.5	67.6	600	5 AAU74597	AAU74597 Antigenic
24	1987.5	66.1	788	4 AAU01903	AAU01903 M. tuberc
25	1960	65.2	744	4 AAU01902	AAU01902 M. tuberc

26	1960	65.2	815	4 AAU01904	AAU01904 M. tuberc
27	1955	65.0	1010	7 ADA26356	Ada26356 Mycobacte
28	1949	64.8	391	2 AAW32381	AAW32381 Mycobacte
29	1949	64.8	391	2 AAW32449	AAW32449 Mycobacte
30	1949	64.8	391	2 AAW64335	AAW64335 Mycobacte
31	1949	64.8	391	2 AAW81702	AAW81702 M. tuberc
32	1949	64.8	391	2 AAY04778	Aay04778 Mycobacte
33	1949	64.8	391	2 AAY38989	Aay38989 M. tuberc
34	1949	64.8	391	2 AAY39132	Aay39132 M. tuberc
35	1949	64.8	391	4 AAU01888	AAU01888 M. tuberc
36	1949	64.8	391	5 AAE29707	Aae29707 Mycobacte
37	1949	64.8	391	5 AAE17571	Aae17571 Mycobacte
38	1949	64.8	394	2 AAY04779	Aay04779 Mycobacte
39	1949	64.8	723	7 ADA26354	Ada26354 Mycobacte
40	1652.5	55.0	396	2 AAW64337	AAW64337 Mycobacte
41	1652.5	55.0	396	2 AAW81704	AAW81704 M. tuberc
42	1652.5	55.0	396	2 AAY38991	Aay38991 M. tuberc
43	1652.5	55.0	396	2 AAY39134	Aay39134 M. tuberc
44	1583	52.6	393	6 ABU36968	ABU36968 Protein e
45	1583	52.6	393	6 ABU34531	ABU34531 Protein e

ALIGNMENTS

RESULT 1
ID AAY32070 standard; protein; 596 AA.

XX AAY32070;

XX AC AAY32070;

XX DT 17-JAN-2000 (first entry)

XX DE Mycobacterium tuberculosis antigen fusion protein Mtb59f.

XX KW Tuberculosis; antigen; fusion protein; Mtb59f; TbH9; Ra35; diagnosis;

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers

FT Peptide 1..8

FT Protein 9..140

FT Protein /note= "Ra12"

FT Protein /note= "TbH9"

XX WO9951748-A2.

XX PD 14-OCT-1999.

XX PF 07-APR-1999; 99WO-US007717.

XX PR 07-APR-1998; 98US-00056556.

XX PR 30-DEC-1998; 98US-00223040.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky YAW, Alderson M, Campos-Neto A;

XX XX WPI; 1999-601610/51.

XX DR N-PSDB; AAZ20205.

XX XX New fusion proteins useful for diagnosis, prevention and treatment of

XX PT tuberculosis.

XX PS Claim 1; Fig 12A-B; 83pp; English.

XX XX This sequence represents a recombinant Mycobacterium tuberculosis bi-

XX CC antigen fusion protein, termed Mtb59f, composed of the antigens TbH9 and

XX CC Ra35. The fusion protein is expressed in host cells using a vector

XX CC carrying a polynucleotide (see AAZ20205) comprising the coding sequences

CC for the 2 antigens. The invention provides fusion proteins (see AAY32059-71) containing at least 2 M. tuberculosis antigens. The new fusion proteins and polynucleotides encoding them are useful as vaccines for preventing tuberculosis (claimed), for diagnosis (via in vitro assays or intradermal skin tests for detection of anti-M. tuberculosis antibodies), monitoring of disease progression, and treatment of tuberculosis. They are more effective immunogens than mixtures of the individual protein components

CC
XX
SQ

Sequence 596 AA;

Query Match 100.0%; Score 3007; DB 2; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.7e-198;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HHHHHHHVDFGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
DB 1 HHHHHHHVDFGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
QY 61 LTVGSWIGSSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPVIA 120
DB 61 LTVGSWIGSSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPVIA 120
QY 121 ENRAELMILITNLGONTPAIYVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEA 180
DB 121 ENRAELMILITNLGONTPAIYVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEA 180
QY 181 PEMTSAGLLLEQAAAEEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLMKTVS 240
DB 181 PEMTSAGLLLEQAAAEEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLMKTVS 240
QY 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
DB 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGP 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGP 360
QY 361 QMLGGLPVQMGARAGGSLGVLVPRPYVPHSPAGDIAPPALSDQDFAPALPLD 420
DB 361 QMLGGLPVQMGARAGGSLGVLVPRPYVPHSPAGDIAPPALSDQDFAPALPLD 420
QY 421 PSAMVAQVGPVAVNTKLGYNNAVAGTGVIPDNGVLTNNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPVAVNTKLGYNNAVAGTGVIPDNGVLTNNHVIAGATDINAFSVGSG 480
QY 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540
DB 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540
QY 541 RVWALGQTVQASDSLTAETLNGLIQDAAIQDSDSGPVPVNGLGQVGMNTAAS 596
DB 541 RVWALGQTVQASDSLTAETLNGLIQDAAIQDSDSGPVPVNGLGQVGMNTAAS 596

RESULT 2
AAE29710
ID AAE29710 standard; protein; 596 AA.
XX
AC AAE29710;
XX
DT 29-AUG-2003 (revised)
DT 27-JAN-2003 (first entry)
XX
XX Mycobacterium sp. MTB59F fusion protein.
DE Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen; Rad5;
KW MTB59F; fusion protein.
XX
XX Mycobacterium sp.
OS Mycobacterium tuberculosis.
OS Chimeric.

XX WO200272792-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
XX (CORI-) CORIXA CORP.
XX
PI Skeiky, Y, Brannon M, Guderian J;
XX
XX WPI; 2002-759944/82.
DR N-PSDB; AAD47086.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.
XX
XX Disclosure; Page 98-99; 155pp; English.
XX
XX The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention are used in methods for eliciting immune response in mammals. They are useful as vaccines to elicit protective immunity against pathogenic microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion polypeptides are used for enhancing the expression of polynucleotides, as in vivo diagnostic agents and for raising antibodies in a non-human animal. The invention is used in gene therapy. The present sequence is MTB59F fusion protein. This fusion protein comprises Rad5 protein from Mycobacterium tuberculosis and TBH9 protein from Mycobacterium sp. (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 596 AA;

Query Match 100.0%; Score 3007; DB 5; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.7e-198;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HHHHHHHVDFGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
DB 1 HHHHHHHVDFGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
QY 61 LTVGSWIGSSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPVIA 120
DB 61 LTVGSWIGSSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPVIA 120
QY 121 ENRAELMILITNLGONTPAIYVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEA 180
DB 121 ENRAELMILITNLGONTPAIYVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEA 180
QY 181 PEMTSAGLLLEQAAAEEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLMKTVS 240
DB 181 PEMTSAGLLLEQAAAEEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLMKTVS 240
QY 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
DB 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGP 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGP 360
QY 361 QMLGGLPVQMGARAGGSLGVLVPRPYVPHSPAGDIAPPALSDQDFAPALPLD 420
DB 361 QMLGGLPVQMGARAGGSLGVLVPRPYVPHSPAGDIAPPALSDQDFAPALPLD 420

QY 421 PSAMVAQGVGVNINIKLGYNNVAGAGTGIVDPNGVLTNNHVITAGATDINAFSVGSG 480
Db 421 PSAMVAQGVGVNINIKLGYNNVAGAGTGIVDPNGVLTNNHVITAGATDINAFSVGSG 480
QY 481 QTYGVVDVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
Db 481 QTYGVVDVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
QY 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLQGVVGMNTAAS 596
Db 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLQGVVGMNTAAS 596

RESULT 3

AAE17574
ID AAE17574 standard; protein; 596 AA.

XX AC AAE17574;
XX AC

DT 22-APR-2002 (first entry)

XX Mycobacterium species MTB59F fusion protein.

XX Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB59F; TbH9-Ra35 protein.

XX Mycobacterium sp.

XX WO200198460-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US019959.

XX 20-JUN-2000; 2000US-00597796.

PR 01-FEB-2001; 2001US-0265737P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Alderson M;
XX WPI; 2002-147798/19.

DR N-PSDB; AAD28344.
XX

Composition comprising MTB39 antigen and MTB32A antigen from
Mycobacterium species, useful for eliciting immune response in a subject.

PS Claim 5; Page 114-115; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC MTB59F (TbH9-Ra35) fusion protein

XX Sequence 596 AA;

99US-00287849.

Query Match 100.0%; Score 3007; DB 5; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.7e-198;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMEHHHHHMFDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLPSAASAFOSVVMG 60

Db 1 HMEHHHHHMFDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLPSAASAFOSVVMG 60

QY 61 LTVGSWIGSSAGLMVAASPYVAMSVTAGAEITAAQVRVAAAAYETAYGLTVPPIVIA 120

Db 61 LTVGSWIGSSAGLMVAASPYVAMSVTAGAEITAAQVRVAAAAYETAYGLTVPPIVIA 120

QY 121 ENRAELMILIAITNLGQNTPAIVNEAEYGEWMAQDAAMFEGYAAATATATATLLPPEEA 180

Db 121 ENRAELMILIAITNLGQNTPAIVNEAEYGEWMAQDAAMFEGYAAATATATATLLPPEEA 180

QY 181 PEMTSAGGLLECAAAVEEASDTAAANQLMNNVPQALQLAQPTGTTTSSKLGGLWKTVS 240

Db 181 PEMTSAGGLLECAAAVEEASDTAAANQLMNNVPQALQLAQPTGTTTSSKLGGLWKTVS 240

QY 241 PHRSPISNMVSMANNHSMNTNSGVMTNTLSSMLKGFAPAAAQAVQTAAGVAMSSSL 300

Db 241 PHRSPISNMVSMANNHSMNTNSGVMTNTLSSMLKGFAPAAAQAVQTAAGVAMSSSL 300

QY 301 GSSLGSSGLGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPG 360

Db 301 GSSLGSSGLGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPG 360

QY 361 QMLGSLPVGMQARAGGSLGVLVPRPYPVNPISPAAGDIAPPALSQDRFADFPALPLD 420

Db 361 QMLGSLPVGMQARAGGSLGVLVPRPYPVNPISPAAGDIAPPALSQDRFADFPALPLD 420

QY 421 PSAMVAQGVGVNINIKLGYNNVAGAGTGIVDPNGVLTNNHVITAGATDINAFSVGSG 480

Db 421 PSAMVAQGVGVNINIKLGYNNVAGAGTGIVDPNGVLTNNHVITAGATDINAFSVGSG 480

QY 481 QTYGVVDVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540

Db 481 QTYGVVDVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLQGVVGMNTAAS 596

Db 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLQGVVGMNTAAS 596

RESULT 4

AAU74599

ID AAU74599 standard; protein; 599 AA.

XX AC AAU74599;

XX 29-AUG-2003 (revised)

DT 08-MAY-2002 (first entry)

XX Antigenic fusion protein Tb59-Ra35 (MtB59f).

XX Fusion protein; tuberculosis; Mycobacterium tuberculosis;
KW tuberculostatic; immunogen; vaccine; Tb59-Ra35; MtB59f.

XX Mycobacterium tuberculosis.

OS Chimeric.

XX Key Location/Qualifiers
FT Misc-difference 597
FT /label= OTHER

FT /note= "OTHER= Xaa. Xaa= In frame stop codon"

XX US2002009459-A1.

XX 24-JAN-2002.

XX 07-APR-1999; 99US-00287849.

13-MAR-1997; 97US-00818112.
01-OCT-1997; 97US-00942578.
18-FEB-1998; 98US-00025197.
07-APR-1998; 98US-00056556.
30-DEC-1998; 98US-00223040.
REED S G.
(SKEI) SKEIKY Y A.
(DILL) DILLON D C.
(ALDE) ALDERSON M.
(CAMP) CAMPOS-NETO A.
Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
WPI; 2002-171134/22.
N-PSDB; ABK14139.
New fusion proteins of Mycobacterium tuberculosis antigens, useful for
diagnosing, treating or preventing M. tuberculosis infection,
particularly as vaccine for treating or preventing tuberculosis.
Claim 1; Fig 12; 62pp; English.
The invention relates to a purified polypeptide which induces an immune
response of Mycobacterium tuberculosis. Polypeptides of the invention are
useful for diagnosing, treating or preventing M. tuberculosis infection,
particularly tuberculosis infection. In particular, the polypeptides are
useful as a vaccine formulation with an adjuvant to afford long-term
protection in animals against the development of tuberculosis. The
protein coding sequence may be used to encode a protein product for use
as an immunogen to induce and/or enhance an immune response to M.
tuberculosis. This sequence represents an M. tuberculosis fusion protein
of the invention. (Updated on 29-AUG-2003 to standardise OS field)
Sequence 599 AA;
Query Match 100.0%; Score 3007; DB 5; Length 599;
Best Local Similarity 100.0%; Pred. No. 1.7e-198;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HMEHHHMHVDFGALPPEINARMYAGPGSLVAAQWDSVASLFAASAFQSVVWG 60
DB 1 HMEHHHMHVDFGALPPEINARMYAGPGSLVAAQWDSVASLFAASAFQSVVWG 60
QY 61 LTGSGWIGSSAGLVAAASPTVAMSVTAGQELTAAQVRAAAAYETAYGLTVPPVIA 120
DB 61 LTGSGWIGSSAGLVAAASPTVAMSVTAGQELTAAQVRAAAAYETAYGLTVPPVIA 120
QY 121 ENRAELMILITNLGQNTPAIYNEAYGEMWAOADAAMFGYAAATATATATLLPPEA 180
DB 121 ENRAELMILITNLGQNTPAIYNEAYGEMWAOADAAMFGYAAATATATATLLPPEA 180
QY 181 PEMSAGLLQAAAVEASDTAAANQIMNNVPAALQQAQPTQGTTPSSKGLGLWKTVS 240
DB 181 PEMSAGLLQAAAVEASDTAAANQIMNNVPAALQQAQPTQGTTPSSKGLGLWKTVS 240
QY 241 PHSPISNMWSMANNHMTSGVSYNTLSSMLKGFAPAAACAVQTAQNGVRAMSSL 300
DB 241 PHSPISNMWSMANNHMTSGVSYNTLSSMLKGFAPAAACAVQTAQNGVRAMSSL 300
QY 301 GSSLGSSGLGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGGP 360
DB 301 GSSLGSSGLGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGGP 360
QY 361 QMLGLPVGQCARAGGSLGVLVPPRPYVPHSPDAAGDIAPALSDQRPAPPLD 420
DB 361 QMLGLPVGQCARAGGSLGVLVPPRPYVPHSPDAAGDIAPALSDQRPAPPLD 420
QY 421 PSAMVAQVGPQVNTKLGYNNAVAGTGVIPDNGVLTNNHVIAGATDINAFVSGS 480
DB 421 PSAMVAQVGPQVNTKLGYNNAVAGTGVIPDNGVLTNNHVIAGATDINAFVSGS 480
QY 481 QTYGVVVGYDRTQDVAVLQLRGAGLPSAALGGVAVGEPVAVMGNSGGGGTTPRAVPG 540

481 QTYGVVVGYDRTQDVAVLQLRGAGLPSAALGGVAVGEPVAVMGNSGGGGTTPRAVPG 540
QY 541 RVVAGQTVQASDSLTGAETINGLIQFDALIQPQDSGPGVNGLGQVVGWNTAAS 596
DB 541 RVVAGQTVQASDSLTGAETINGLIQFDALIQPQDSGPGVNGLGQVVGWNTAAS 596
RESULT 5
AAO22142
ID AAO22142 standard; protein; 729 AA.
XX AC AAO22142;
XX 03-OCT-2002 (first entry)
XX DE Ra12-H9-32A fusion protein.
XX Ra12; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
KW vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
KW immunogen; cytokine.
XX Unidentified.
OS Chimeric.
XX WO200125401-A2.
XX 12-APR-2001.
XX 06-OCT-2000; 2000WO-US027652.
XX 07-OCT-1999; 99US-0158583P.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Guderian J;
XX WPI; 2001-266299/27.
XX N-PSDB; AAL40773.
XX Recombinant nucleic acid molecule for producing high yield expression of
desired fusion polypeptides, encodes fusion polypeptide comprising
Mycobacterium tuberculosis coding sequence and heterologous polypeptide.
XX Disclosure; Fig 6; 39pp; English.
XX The invention relates to a recombinant nucleic acid molecule encoding a
fusion polypeptide, comprising a polynucleotide sequence of Ra12, a 14
kDa C-terminal fragment of serine protease antigen MTB32A of
Mycobacterium tuberculosis, and a heterologous polynucleotide sequence.
XX The recombinant fusion nucleic acids and polypeptides are useful for
providing stable and high yield expression of fusion polypeptides of both
eukaryotic and prokaryotic origin and to encode a protein product for use
as an antigen for detecting serum antibodies. The presence of serum
antibodies to M. tuberculosis antigens in an individual indicates that
the individual is infected with it. The fusion polypeptides are useful as
sources of proteins for monitoring binding of serum antibodies to fusion
proteins and as an immunogen to induce and/or enhance immune responses.
XX the coding sequences can be ligated with a coding sequence of another
molecule such as M. tuberculosis antigens, a cytokine or an adjuvant and
can be used in vivo as a DNA vaccine. This sequence represents the Ra12-
H9-32A fusion protein
XX Sequence 729 AA;
Query Match 98.0%; Score 2946; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.4e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MYDFGALPPEINARMYAGPGSLVAAQWDSVASLFAASAFQSVVWGLTVGSMIG 68
DB 142 MYDFGALPPEINARMYAGPGSLVAAQWDSVASLFAASAFQSVVWGLTVGSMIG 201

XX WO2003070187-A2.
FN
PD 28-AUG-2003.
XX
PF 18-FEB-2003; 2003WO-US004903.
XX
PR 15-FEB-2002; 2002US-0357351P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Guderian J, Reed S;
XX
DR WPI; 2003-697554/66.
XX N-PSDB; ADA26359.
DR
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
XX and MTB39 antigens, with or without the MTB85A antigen, from a
XX Mycobacterium species, useful for diagnosing, preventing and/or treating
XX tuberculosis infection.
XX
PS Claim 85; Fig 14; 112pp; English.
XX
XX The invention relates to a novel nucleic acid encoding a fusion
XX polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
XX MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
XX A polypeptide of the invention has tuberculostatic activity. A
XX polynucleotide of the invention may have a use in gene therapy, and as a
XX vaccine. The methods and compositions of the present invention are useful
XX for diagnosing, preventing and/or treating tuberculosis infection. The
XX present sequence is used in the exemplification of the invention.
XX
SQ Sequence 825 AA;

Query Match 98.0%; Score 2946; DB 7; Length 825;
Best Local Similarity 100.0%; Pred. No. 4e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMWIG 68
DB 142 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMWIG 201

QY 69 SSAGLWVAASFPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
DB 202 SSAGLWVAASFPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 261

QY 129 LIATNLLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATATATLPPFEAPEMTSAGG 188
DB 262 LIATNLLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 321

QY 189 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQTPQTTPSSKLGGLWKTVPSPHRSPI 248
DB 322 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQTPQTTPSSKLGGLWKTVPSPHRSPI 381

QY 249 MYSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQVAMSLSSLSGSSG 308
DB 382 MYSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQVAMSLSSLSGSSG 441

QY 309 LGGVAAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAAEFGPGQMLGLPV 368
DB 442 LGGVAAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAAEFGPGQMLGLPV 501

QY 369 GQMGARAGGSLGVLVPPRPYVMPHSPAAGDIAPPALSQDRPADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGSLGVLVPPRPYVMPHSPAAGDIAPPALSQDRPADFPALPLDPSAMVAQV 561

QY 429 GPQVNVINTKLGNNAVCAGTGVIDPENGVLVTHNVHAGATDINAFSGSGQTYGVDDV 498
DB 562 GPQVNVINTKLGNNAVCAGTGVIDPENGVLVTHNVHAGATDINAFSGSGQTYGVDDV 621

QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPEPVVAMGNSGGQGTTPRAVGRVVALGQT 548
DB 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPEPVVAMGNSGGQGTTPRAVGRVVALGQT 681

QY 549 VOASDSLTCABETLNGLIQFDAAIQPGSDSGPVPVNGLGQVYGMNTAAS 596
DB 682 VOASDSLTCABETLNGLIQFDAAIQPGSDSGPVPVNGLGQVYGMNTAAS 729

RESULT 8
ADA26365
ID ADA26365 standard; protein; 875 AA.
XX
AC ADA26365;
XX
DT 20-NOV-2003 (first entry)
XX
DE Mycobacterium MTB72F-Erd14 (fusion MTB89F) protein.
XX
KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
OS Chimeric.
OS Mycobacterium sp.
XX
PN WO2003070187-A2.
XX
PD 28-AUG-2003.
XX
PF 18-FEB-2003; 2003WO-US004903.
XX
PR 15-FEB-2002; 2002US-0357351P.
XX
PA (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Guderian J, Reed S;
XX
XX WPI; 2003-697554/66.
XX N-PSDB; ADA26358.
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
XX and MTB39 antigens, with or without the MTB85A antigen, from a
XX Mycobacterium species, useful for diagnosing, preventing and/or treating
XX tuberculosis infection.
XX
XX Claim 85; Fig 13; 112pp; English.
XX
XX The invention relates to a novel nucleic acid encoding a fusion
XX polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
XX MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
XX A polypeptide of the invention has tuberculostatic activity. A
XX polynucleotide of the invention may have a use in gene therapy, and as a
XX vaccine. The methods and compositions of the present invention are useful
XX for diagnosing, preventing and/or treating tuberculosis infection. The
XX present sequence is used in the exemplification of the invention.
XX
SQ Sequence 875 AA;

Query Match 98.0%; Score 2946; DB 7; Length 875;
Best Local Similarity 100.0%; Pred. No. 4.3e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMWIG 68
DB 142 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMWIG 201

QY 69 SSAGLWVAASFPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
DB 202 SSAGLWVAASFPYVAMSVTGAQELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 261

QY 129 LIATNLLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 188
DB 262 LIATNLLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPEMTSAGG 321

QY 189 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQTPQTTPSSKLGGLWKTVPSPHRSPI 248
DB 322 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQTPQTTPSSKLGGLWKTVPSPHRSPI 381

Db 322 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQCTPSSKLGGLWKTVPSPHRSPI 381
QY 249 MVSMAHHMSTNSGVSMTNTLSMLKGFAPAAAAQVQTAQNGVRAMSSLGSSLGSSG 308
Db 382 MVSMAHHMSTNSGVSMTNTLSMLKGFAPAAAAQVQTAQNGVRAMSSLGSSLGSSG 441
QY 309 LGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 368
Db 442 LGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 501
QY 369 GQMGARAGGSLGVLRVPPRPYMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRVPPRPYMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNNINTKLYNNVAVGAGTGVIDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 488
Db 562 GPQVNNINTKLYNNVAVGAGTGVIDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 621
QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMNSGGCGGTPRAVPGRVVAGQT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMNSGGCGGTPRAVPGRVVAGQT 681
QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS 596
Db 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS 729
RESULT 9
AAE29731
ID AAE29731 standard; protein; 930 AA.
XX
AC AAE29731;
XX
DT 29-AUG-2003 (revised)
DT 27-JAN-2003 (first entry)
DE
DE Mycobacterium sp. MTB72F-Leishmania sp. MAPS (aka r95f) fusion protein.
XX
KW Vaccine; immunity; diagnostic agent; gene therapy; MTB72F; MAPS;
KW fusion protein.
XX
OS Mycobacterium sp.
OS Leishmania sp.
OS Chimeric.
XX
FN WO20027292-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Brannon M, Guderian J;
PI
XX NFI; 2002-759844/82.
DR N-PSDB; AAD47110.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif,
PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
PT tuberculosis.
XX
XX Example 6; Page 129-132; 155pp; English.
PS
XX The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention

CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC Mycobacterium sp. MTB72F-Leishmania sp. thiol specific antitoxin [TSA;
CC MAPS (aka r95f)] fusion protein. This sequence comprises Mycobacterium
CC species MTB72F (a 72 kDa poly-protein fusion construct comprising Ral2-
CC TbH9-Ra35) linked to the Leishmania TSA (MAPS). (Updated on 29-AUG-2003
CC to standardise OS field)
XX
SQ Sequence 930 AA;
Query Match 98.0%; Score 2946; DB 5; Length 930;
Best Local Similarity 100.0%; Pred. No. 4,6e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDFGALPPEINSAARMYAGPGSASLVAAAQMWDSVADLSFSAASAFQSVVWGLTVGSMIG 68
Db 142 MVDFGALPPEINSAARMYAGPGSASLVAAAQMWDSVADLSFSAASAFQSVVWGLTVGSMIG 201
QY 69 SSAGLWVAASPYVAMSVTAGQAELETAQVRAAAAYETAYGLTVPDPVIAEARELMI 128
Db 202 SSAGLWVAASPYVAMSVTAGQAELETAQVRAAAAYETAYGLTVPDPVIAEARELMI 261
QY 129 LIATNLLGONTPTAIAVNEAEYGEWMAQDAANFGYAAATATATATLLPFEAPEMTSAGG 188
Db 262 LIATNLLGONTPTAIAVNEAEYGEWMAQDAANFGYAAATATATATLLPFEAPEMTSAGG 321
QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQCTPSSKLGGLWKTVPSPHRSPI 248
Db 322 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQCTPSSKLGGLWKTVPSPHRSPI 381
QY 249 MVSMAHHMSTNSGVSMTNTLSMLKGFAPAAAAQVQTAQNGVRAMSSLGSSLGSSG 308
Db 382 MVSMAHHMSTNSGVSMTNTLSMLKGFAPAAAAQVQTAQNGVRAMSSLGSSLGSSG 441
QY 309 LGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 368
Db 442 LGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 501
QY 369 GQMGARAGGSLGVLRVPPRPYMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRVPPRPYMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNNINTKLYNNVAVGAGTGVIDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 488
Db 562 GPQVNNINTKLYNNVAVGAGTGVIDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 621
QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMNSGGCGGTPRAVPGRVVAGQT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVWAMNSGGCGGTPRAVPGRVVAGQT 681
QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS 596
Db 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS 729
RESULT 10
ADA26364
ID ADA26364 standard; protein; 930 AA.
XX
AC ADA26364;
XX
DT 20-NOV-2003 (first entry)
XX
DE Mycobacterium MTB72F-MAPS (fusion r95F) protein.
KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
OS Chimeric.

OS Mycobacterium sp.
 XX WO2003070187-A2.
 XX 28-AUG-2003.
 XX 18-FEB-2003; 2003WO-US004903.
 XX 15-FEB-2002; 2002US-0357351P.
 PR (CORI-) CORIXA CORP.
 PA Skeiky Y, Guderian J, Reed S;
 XX WPI; 2003-697554/66.
 XX N-PSDB; ADA26357.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX Disclosure; Fig 12; 112pp; English.
 PS The invention relates to a novel nucleic acid encoding a fusion
 XX polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX Sequence 930 AA;
 SQ

Query Match 98.0%; Score 2946; DB 7; Length 930;
 Best Local Similarity 100.0%; Pred. No. 4.6e-194;
 Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
 DB 142 MVDGALPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLWMAAASPYYAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
 DB 202 SSAGLWMAAASPYYAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 261

QY 129 LIATNLLQONTPTAIANVEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPBMTSAGG 188
 DB 262 LIATNLLQONTPTAIANVEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPBMTSAGG 321

QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248
 DB 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 381

QY 249 MYSNANNHMTNSGVSMTNTLSNMLKGFAPAPAAAQVQTAQNGVRAMSLGSSLCSSG 308
 DB 382 MYSNANNHMTNSGVSMTNTLSNMLKGFAPAPAAAQVQTAQNGVRAMSLGSSLCSSG 441

QY 309 LGGGVAANLGRAASVGSLSVQAAANAQVTPAARALPLTSLTSAERGGQMLGLPV 368
 DB 442 LGGGVAANLGRAASVGSLSVQAAANAQVTPAARALPLTSLTSAERGGQMLGLPV 501

QY 369 QMGARAGGSLGVLPRPPRYVPHSPAGDIAAPPALSDQRFADFPALPLDPSAMVAQV 428
 DB 502 QMGARAGGSLGVLPRPPRYVPHSPAGDIAAPPALSDQRFADFPALPLDPSAMVAQV 561

QY 429 GPQVYNITKLGYNNAGVAGTIVIDPQVVLNNHVIAGATDINAFVSGGQTYGVDVV 488
 DB 562 GPQVYNITKLGYNNAGVAGTIVIDPQVVLNNHVIAGATDINAFVSGGQTYGVDVV 621

QY 489 GYDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPPYVAMGNSGGQGTTPRAVPRVVALQGT 548

DB 622 GYDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPPYVAMGNSGGQGTTPRAVPRVVALQGT 681

QY 549 VOASDSLTCGAETLNGLIQFDAAIQFGDSGGFPVNVNGLGQVVGMTAAS 596
 DB 682 VOASDSLTCGAETLNGLIQFDAAIQFGDSGGFPVNVNGLGQVVGMTAAS 729

RESULT 11
 ADA26370
 ID ADA26370 standard; protein; 1016 AA.
 XX ADA26370;
 XX 20-NOV-2003 (first entry)
 XX M. bovis MTB72F and 85b complex antigen (fusion MTB103F).
 XX fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine; MTB72F; 85B.
 XX Chimeric.
 OS Mycobacterium bovis.
 XX WO2003070187-A2.
 XX 28-AUG-2003.
 XX 18-FEB-2003; 2003WO-US004903.
 XX 15-FEB-2002; 2002US-0357351P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Guderian J, Reed S;
 XX WPI; 2003-697554/66.
 XX N-PSDB; ADA26363.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX Claim 85; Fig 18; 112pp; English.
 PS The invention relates to a novel nucleic acid encoding a fusion
 XX polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX Sequence 1016 AA;
 SQ

Query Match 98.0%; Score 2946; DB 7; Length 1016;
 Best Local Similarity 100.0%; Pred. No. 5.1e-194;
 Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
 DB 142 MVDGALPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLWMAAASPYYAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
 DB 202 SSAGLWMAAASPYYAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 261

QY 129 LIATNLLQONTPTAIANVEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 188
 DB 262 LIATNLLQONTPTAIANVEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 321

QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248

Db 322 LLEQAAVEEASDTAAANQNNVPAALQQLAQPTQGTTFSSKLGKLVKTVSPHRSPISN 381
QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSSLGSSIGSSG 308
Db 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSSLGSSIGSSG 441
QY 309 LGGVVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAARPGQMLGGLPV 368
Db 442 LGGVVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAARPGQMLGGLPV 501
QY 369 GQMGARAGGSLGVLRPPRYPMHSPAAAGDIAPPALSDRFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRPPRYPMHSPAAAGDIAPPALSDRFPALPLDPSAMVAQV 561
QY 429 GPQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVV 488
Db 562 GPQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVV 621
QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGNSSGGQGTTPRAVPGRVVVALGQT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGNSSGGQGTTPRAVPGRVVVALGQT 681
QY 549 VOASDSLTAETLNGLIQFDAAIQPDGSGGPVVGNGIGOVVGMNTAAS 596
Db 682 VOASDSLTAETLNGLIQFDAAIQPDGSGGPVVGNGIGOVVGMNTAAS 729

RESULT 12

ADA26369
ID ADA26369 standard; protein; 1022 AA.

XX AC ADA26369;
XX DT 20-NOV-2003 (first entry)
XX DE Mycobacterium MTB72F-mTCC#1 (fusion MTB102tm2P) protein.

XX KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
XX KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX OS Chimeric.
XX OS Mycobacterium sp.

XX PN WO2003070187-A2.
XX PD 28-AUG-2003.

XX PF 18-FEB-2003; 2003WO-US004903.
XX PR 15-FEB-2002; 2002US-0357351P.

XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Guderian J, Reed S;
XX PR WPI; 2003-697554/66.
XX DR N-PSDB; ADA26362.

XX XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX XX Claim 85; Fig 17; 112pp; English.
PS The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention, are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The

CC present sequence is used in the exemplification of the invention.

XX SQ Sequence 1022 AA;

Query Match 99.0%; Score 2946; DB 7; Length 1022;
Best Local Similarity 100.0%; Pred. No. 5.2e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDFGALPPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFOSVVMGLTVGWSIG 68
Db 142 MVDFGALPPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFOSVVMGLTVGWSIG 201

QY 59 SSAGLWVAASPYVWMSVITAGQAELETAQVRVAAAAYETAYGLTVPPVTAENRAELMI 128
Db 202 SSAGLWVAASPYVWMSVITAGQAELETAQVRVAAAAYETAYGLTVPPVTAENRAELMI 261

QY 129 LIATNLLGONTPALAVNEAEYGEWMAQDAANFVAAAATATATATLPPFEAPEMTSAGG 188
Db 262 LIATNLLGONTPALAVNEAEYGEWMAQDAANFVAAAATATATATLPPFEAPEMTSAGG 321

QY 189 LLEQAAVEEASDTAAANQNNVPAALQQLAQPTQGTTFSSKLGKLVKTVSPHRSPISN 248
Db 322 LLEQAAVEEASDTAAANQNNVPAALQQLAQPTQGTTFSSKLGKLVKTVSPHRSPISN 381

QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSSLGSSIGSSG 308
Db 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSSLGSSIGSSG 441

QY 309 LGGVVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAARPGQMLGGLPV 368
Db 442 LGGVVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAARPGQMLGGLPV 501

QY 369 GQMGARAGGSLGVLRPPRYPMHSPAAAGDIAPPALSDRFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRPPRYPMHSPAAAGDIAPPALSDRFPALPLDPSAMVAQV 561

QY 429 GPQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVV 488
Db 562 GPQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVV 621

QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGNSSGGQGTTPRAVPGRVVVALGQT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGNSSGGQGTTPRAVPGRVVVALGQT 681

QY 549 VOASDSLTAETLNGLIQFDAAIQPDGSGGPVVGNGIGOVVGMNTAAS 596
Db 682 VOASDSLTAETLNGLIQFDAAIQPDGSGGPVVGNGIGOVVGMNTAAS 729

RESULT 13

ADA26368
ID ADA26368 standard; protein; 1154 AA.

XX AC ADA26368;
XX DT 20-NOV-2003 (first entry)

XX DE Mycobacterium MTB72F-mTCC#2 (fusion MTB114F) protein.
XX KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
XX KW tuberculosis; tuberculostatic; gene therapy; vaccine.

XX OS Chimeric.
XX OS Mycobacterium sp.

XX PN WO2003070187-A2.
XX PD 28-AUG-2003.

XX PF 18-FEB-2003; 2003WO-US004903.
XX PR 15-FEB-2002; 2002US-0357351P.

PA (CORI-) CORIXA CORP.
 XX Skeiky Y, Guderman J, Reed S;
 XX WPI: 2003-697554/66.
 DR N-PSDB; ADA26361.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX Claim 85; Fig 16; 112pp; English.
 XX The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX Sequence 1154 AA;
 SQ
 Query Match 98.0%; Score 2946; DB 7; Length 1154;
 Best Local Similarity 100.0%; Pred. No. 6e-194;
 Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
 Db 142 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201
 QY 69 SSAGLWAAAAPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 Db 202 SSAGLWAAAAPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
 QY 129 LIATNLGQNTPAIVNEAEYGEWQAQAAAFVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 188
 Db 262 LIATNLGQNTPAIVNEAEYGEWQAQAAAFVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 321
 QY 249 MVSNNHMSMTNGSVNTLSSMLKGFAPAAAQAVQTAQNGVRAMSLSGLSSG 308
 Db 382 MVSNNHMSMTNGSVNTLSSMLKGFAPAAAQAVQTAQNGVRAMSLSGLSSG 441
 QY 309 LGGVAAALGRAASVGSLSVQAAWAAANQAVTPAARALPLTSLTSAERGPQWGLGLPV 368
 Db 442 LGGVAAALGRAASVGSLSVQAAWAAANQAVTPAARALPLTSLTSAERGPQWGLGLPV 501
 QY 369 GQMGARAGGSLGVLVPPRYVPHSPAGDIAPPALSOORFADFPALPLDPSAMVAQV 428
 Db 502 GQMGARAGGSLGVLVPPRYVPHSPAGDIAPPALSOORFADFPALPLDPSAMVAQV 561
 QY 429 GPVVNINIKLYNNAVGAGTGVIDPVGVLVTHNHVIAAGTIDINAFSVSGGQTYGVDDV 488
 Db 562 GPVVNINIKLYNNAVGAGTGVIDPVGVLVTHNHVIAAGTIDINAFSVSGGQTYGVDDV 621
 QY 489 GYDRTQDVAVLQLRGAGLPSAATGGGVAVGCEPVVAMGNSGGQGTTPRVPGRVVALGQT 548
 Db 622 GYDRTQDVAVLQLRGAGLPSAATGGGVAVGCEPVVAMGNSGGQGTTPRVPGRVVALGQT 681
 QY 549 VQASDSLTCAGETLNLGIQDAAIQPGDSGSPVNGLGQVVGWNTAAS 596
 Db 682 VQASDSLTCAGETLNLGIQDAAIQPGDSGSPVNGLGQVVGWNTAAS 729
 RESULT 14
 AAE29709
 ID AAE29709 standard; protein; 729 AA.

XX AAE29709;
 XX 29-AUG-2003 (revised)
 XX 27-JAN-2003 (first entry)
 XX Mycobacterium sp. MTB72FmutSA fusion protein.
 XX Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen;
 XX Ra35MutSA; Ra12; MTB72MutSA; fusion protein.
 XX Mycobacterium sp.
 XX Mycobacterium tuberculosis.
 XX Chimeric.
 XX WO200272792-A2.
 XX 19-SEP-2002.
 XX 13-MAR-2002; 2002WO-US008223.
 XX 13-MAR-2001; 2001US-0275837P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Brannon M, Guderman J;
 XX WPI: 2002-759844/82.
 XX N-PSDB; AAD47084.
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX Disclosure; Page 93-95; 155pp; English.
 XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC MTB72F fusion protein. This fusion protein comprises Ra35MutSA mutant
 CC protein and Ra12 protein from M. tuberculosis and TBH9 protein from
 CC Mycobacterium sp. (Updated on 29-AUG-2003 to standardise OS field)
 XX Sequence 729 AA;
 SQ
 Query Match 97.9%; Score 2943; DB 5; Length 729;
 Best Local Similarity 99.8%; Pred. No. 5.4e-194;
 Matches 587; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 9 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
 Db 142 MVDGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201
 QY 69 SSAGLWAAAAPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 Db 202 SSAGLWAAAAPYVAVMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
 QY 129 LIATNLGQNTPAIVNEAEYGEWQAQAAAFVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 188
 Db 262 LIATNLGQNTPAIVNEAEYGEWQAQAAAFVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 321
 QY 189 LLEQAAAVEEASDTAAANQLMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPTSN 248

Db 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381
 Qy 249 MYSMANNHMSMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVYRAMSSLGSSLGSSG 308
 Db 382 MYSMANNHMSMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVYRAMSSLGSSLGSSG 441
 Qy 309 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGMGLGLPV 368
 Db 442 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGMGLGLPV 501
 Qy 369 GQMGARAGGSLGVLRVPPRYPMHPSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 428
 Db 502 GQMGARAGGSLGVLRVPPRYPMHPSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 561
 Qy 429 GPQWNINTKLGYNNAVAGCTGVIDPENGVLNNHVIAGATDINAFSGSGGTGYGVDV 488
 Db 562 GPQWNINTKLGYNNAVAGCTGVIDPENGVLNNHVIAGATDINAFSGSGGTGYGVDV 521
 Qy 489 GYDRTQDVAVLQLRGAGGLPSAAGISVAGGVAVGEPVWAMGSGGGTTPRAVPGRVVAGT 548
 Db 622 GYDRTQDVAVLQLRGAGGLPSAAGISVAGGVAVGEPVWAMGSGGGTTPRAVPGRVVAGT 681
 Qy 549 VQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQVGMNTAAS 596
 Db 682 VQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQVGMNTAAS 729

RESULT 15

AAE17573
 ID AAE17573 standard; protein; 729 AA.

AC AAE17573;

XX 22-APR-2002 (first entry)

XX Mycobacterium species MTB72FmutSA fusion protein.

XX Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccines; MTB72FmutSA; Ra12-TbH9-Ra35MutSA;
 KW mutant; mutein.

XX Mycobacterium sp.

XX Key Location/Qualifiers
 FH Region 8..139
 FT /label= Ra12_protein
 FT Region 142..532
 FT /label= TbH9FL_protein
 FT Region 535..729
 FT /label= Ra35_protein
 FT Misc-difference 710
 FT /note= "Wild type Ser substituted with Ala"

XX WO200198460-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US019959.

XX 20-JUN-2000; 2000US-00597796.

XX 01-FEB-2001; 2001US-0265737P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Alderson M;

XX WPI; 2002-147798/19.

XX N-PSDB; AAD28343.

XX Composition comprising MTB39 antigen and MTB32A antigen from

XX Mycobacterium species, useful for eliciting immune response in a subject.

XX Claim 88; Fig 5; 136pp; English.

XX

CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB72FmutSA (Ra12-TbH9-Ra35MutSA) mutant fusion protein

XX Sequence 729 AA;

Query Match 97.9%; Score 2943; DB 5; Length 729;

Best Local Similarity 99.8%; Pred. No. 5.4e-194;

Matches 587; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 9 MVDGALPPEINSARMYAGPGSASIVAAQAQWDSVASDLFSAASAFQSVVGLTVGSWIG 68

Db 142 MVDGALPPEINSARMYAGPGSASIVAAQAQWDSVASDLFSAASAFQSVVGLTVGSWIG 201

Qy 69 SSAGLMVAASPFYVAMSVTAGQAELETAQAQVRAAAAYETAYGLTVPPIAENRAELMI 128

Db 202 SSAGLMVAASPFYVAMSVTAGQAELETAQAQVRAAAAYETAYGLTVPPIAENRAELMI 261

Qy 129 LIATNLIQNTPATAVNEAEGEMWQAQDAAMFGYAAATATATATLPPPEAPEMTSAGG 188

Db 262 LIATNLIQNTPATAVNEAEGEMWQAQDAAMFGYAAATATATATLPPPEAPEMTSAGG 321

Qy 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248

Db 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381

Qy 249 MYSMANNHMSMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVYRAMSSLGSSLGSSG 308

Db 382 MYSMANNHMSMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVYRAMSSLGSSLGSSG 441

Qy 309 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGMGLGLPV 368

Db 442 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGMGLGLPV 501

Qy 369 GQMGARAGGSLGVLRVPPRYPMHPSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 428

Db 502 GQMGARAGGSLGVLRVPPRYPMHPSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 561

Qy 429 GPQWNINTKLGYNNAVAGCTGVIDPENGVLNNHVIAGATDINAFSGSGGTGYGVDV 488

Db 562 GPQWNINTKLGYNNAVAGCTGVIDPENGVLNNHVIAGATDINAFSGSGGTGYGVDV 621

Qy 489 GYDRTQDVAVLQLRGAGGLPSAAGISVAGGVAVGEPVWAMGSGGGTTPRAVPGRVVAGT 548

Db 622 GYDRTQDVAVLQLRGAGGLPSAAGISVAGGVAVGEPVWAMGSGGGTTPRAVPGRVVAGT 681

Qy 549 VQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQVGMNTAAS 596

Db 682 VQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQVGMNTAAS 729

Search completed: June 30, 2004, 16:48:49

JOB time : 63.2853 secs

